CENGAGE Learning

ESTADÍSTICA MATEMÁTICA

CON APLICACIONES

SÉPTIMA EDICIÓN

WACKERLY • MENDENHALL • SCHEAFFER

SÉPTIMA EDICIÓN

Estadística matemática con aplicaciones

Dennis D. Wackerly
University of Florida

William Mendenhall III
University of Florida, Emeritus

Richard L. Scheaffer
University of Florida, Emeritus

Traducción:
Jorge Humberto Romo Muñoz
Traductor profesional

Revisión técnica:
Dra. Ana Elizabeth García Hernández
Universidad La Salle Morelia

CENGAGE Learning™

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

Estadística matemática con aplicaciones

Estadística matemática con aplicaciones
Séptima Edición
Wackerly, Dennis D./
William Mendenhall III/
Richard L. Scheaffer

Presidente de Cengage Learning Latinoamérica:
Javier Arellano Gutiérrez

Director general México y Centroamérica:
Pedro Turbay Garrido

Director editorial Latinoamérica:
José Tomás Pérez Bonilla

Director de producción:
Raúl D. Zendejas Espejel

Coordinadora editorial:
María Rosas López

Editor:
Sergio R. Cervantes González

Editor de producción:
Timoteo Eliosa García

Ilustrador:
Erik Adigard, Patricia McShane

Diseño de portada:
Grupo Insigne OTA

Composición tipográfica:
Ediciones OVA

Traducido del libro:
Mathematical statistics with applications, 7th ed.
Wackerly, Dennis D./William Mendenhall III/
Richard L. Scheaffer
Publicado en inglés por Thomson/Brooks-Cole, © 2008
ISBN-13: 978-0-495-11081-1
ISBN-10: 0-495-11081-7

Datos para catalogación bibliográfica:
Estadística matemática con aplicaciones
Séptima Edición
Wackerly, Dennis D./William Mendenhall III/
Richard L. Scheaffer
ISBN-13: 978-607-481-399-9
ISBN-10: 607-481-399-X

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

CONTENIDO

PREFACIO

El propósito y requisitos previos de este libro

Estadística matemática con aplicaciones fue escrito para el uso de los estudiantes en cursos universitarios de un año (9 horas-trimestre o 6 horas-semestre) de estadística matemática. El propósito de este libro es presentar una sólida base al estudiante, de la teoría estadística y al mismo tiempo darle orientación de la relevancia e importancia de la teoría para resolver problemas prácticos del mundo real. Pensamos que un curso de este tipo es apropiado para casi todas las disciplinas de licenciatura, incluyendo matemáticas, donde el contacto con aplicaciones puede ser una experiencia motivadora y alentadora. El único requisito matemático previo es un conocimiento profundo de cálculo universitario de primer año, incluyendo sumas de series infinitas, derivadas e integración simple y doble.

Nuestro método

El hablar con estudiantes que estén tomando un curso inicial de estadística matemática, o lo hayan terminado, reveló una falla importante en numerosos cursos. Los estudiantes pueden tomar el curso y terminarlo sin entender claramente la naturaleza de la estadística. Muchos de ellos ven la teoría como un conjunto de temas, con una relación débil o fuerte entre ellos, pero no ven que la estadística es una teoría de información con inferencia como su meta. Además, pueden terminar el curso sin entender el importante papel que desempeñan las estadísticas en investigaciones científicas.

Estas consideraciones nos llevaron a crear un texto que difiere de otros en tres formas:

- Primero, la presentación de la probabilidad está precedida de un claro enunciado del objetivo de la estadística —*la inferencia estadística*— y su función en la investigación científica. A medida que los estudiantes avanzan en la teoría de probabilidades (Capítulos 2 al 7), se les recuerda con frecuencia el papel que los temas principales desempeñan en la inferencia estadística. El efecto acumulativo es que la inferencia estadística es el tema dominante del curso.
- La segunda característica del texto es su *conectividad*. Explicamos no sólo la forma en que los temas principales desempeñan un papel en inferencia estadística, sino que también cómo es que los temas están relacionados entre sí. Estas interesantes explicaciones aparecen con más frecuencia en las introducciones y conclusiones del capítulo.
- Por último, el texto es único en su *énfasis práctico*, en ejercicios en todo el libro y en los útiles temas metodológicos estadísticos contenidos en los Capítulos del 11 al

15, cuya meta es reforzar la base elemental pero sólida desarrollada en los capítulos iniciales.

El libro se puede usar en varias formas y adaptarse a los gustos de estudiantes y profesores. La dificultad del material puede aumentar o disminuir si se controla la asignación de los ejercicios, se eliminan algunos temas y si se hace variar el tiempo dedicado a cada tema. Se puede dar un toque de más aplicación al eliminar algunos temas, por ejemplo algunas secciones de los Capítulos 6 y 7, y dedicar más tiempo a los capítulos aplicados al final.

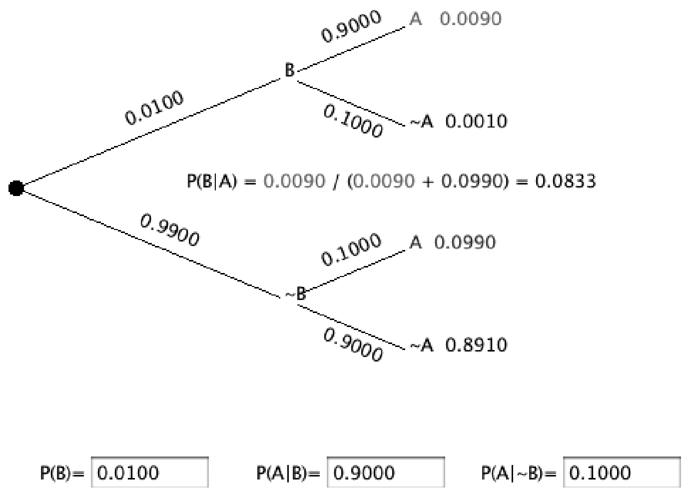
Cambios en la séptima edición

Existe una gran cantidad de estudiantes que aprovechan la información visual de conceptos y resultados. Algo nuevo en esta edición es la inclusión de aplicaciones breves de computadora, todas ellas disponibles en línea para su uso en el sitio web Cengage, www.cengage.com/statistics/wackerly. Algunas de estas aplicaciones se utilizan para demostrar conceptos estadísticos, otras permiten a los usuarios evaluar el impacto de selecciones de parámetro en la forma de funciones de densidad y el resto de aplicaciones se pueden usar para hallar probabilidades exactas y cuantiles asociados con variables aleatorias gamma, beta, normal, χ^2, t y F distribuidas, todo ello información importante cuando se construyen intervalos de confianza o se realizan pruebas de hipótesis. Algunas de las aplicaciones aportan información existente por medio del uso de otros paquetes de software. Destacan entre éstas, el lenguaje R y el ambiente para computación y gráficas estadísticas (disponibles gratis en http://www.r-project.org/) se puede usar para obtener los cuantiles y probabilidades asociados con las distribuciones discretas y continuas ya antes citadas. Los comandos R apropiados se presentan en las secciones respectivas de los Capítulos 3 y 4. La ventaja de las aplicaciones breves de computadora es que son de estilo "apunte y dispare", dan gráficas adjuntas y son considerablemente más fáciles de usar. No obstante, R es mucho más potente que estas aplicaciones y se puede usar para muchos otros fines estadísticos. Dejamos otras aplicaciones de R al usuario interesado o profesor.

El Capítulo 2 introduce la primera aplicación breve (applet), *Bayes' Rule as a Tree* (*la Regla de Bayes como Árbol*), demostración que permite a los usuarios ver por qué a veces se obtienen resultados sorprendentes cuando se aplica la regla de Bayes (vea Figura 1). Al igual que en la sexta edición, las estimaciones de máxima verosimilitud se introducen en el Capítulo 3 mediante ejemplos para las estimaciones de los parámetros de las distribuciones binomiales, geométricas y binomiales negativas basadas en valores numéricos específicos observados de variables aleatorias que poseen estas distribuciones. Los problemas de seguimiento al final de las secciones respectivas se amplían en estos ejemplos.

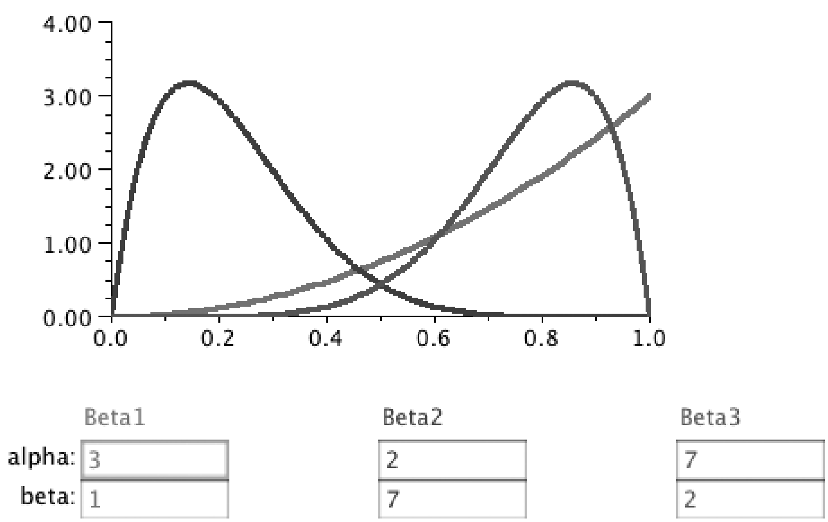
En el Capítulo 4, la aplicación *Normal Probabilities* (*Probabilidades Normales)* se usa para calcular la probabilidad de que cualquier variable aleatoria normalmente distribuida, especificada por el usuario, caiga en cualquier intervalo especificado. También proporciona una gráfica de la función de densidad normal seleccionada y un refuerzo visual del hecho de que las probabilidades asociadas con cualquier variable aleatoria normalmente distribuida son

FIGURA 1
Ilustración de una aplicación breve de la regla de Bayes



equivalentes a probabilidades asociadas con la distribución normal estándar. La aplicación *Normal Probabilities (One Tail)* (*Probabilidades Normales (Una Cola))* proporciona áreas de cola superior asociadas con cualquier distribución normal, especificada por el usuario y también se puede usar para establecer el valor que corta un área, especificada por un usuario, en la cola superior para cualquier variable aleatoria normalmente distribuida. Las probabilidades y cuantiles asociados con variables aleatorias normales estándar se obtienen al seleccionar la media de valores de parámetro = 0 y desviación estándar = 1. Las distribuciones beta y gamma se exploran más a fondo en este capítulo. Los usuarios pueden al mismo tiempo graficar tres densidades gamma (o beta) (todas con valores de parámetros seleccionados por el usuario) y evaluar el impacto que los valores de parámetro tienen en las formas de funciones de densidad gamma (o beta) (vea Figura 2). Esto se logra con el uso de las aplicaciones breves *Comparison*

FIGURA 2
Comparación de aplicación breve de tres densidades beta



of Gamma Density Functions and *Comparison of Beta Density Functions,* (*Comparación de Funciones de Densidad Gamma* y *Comparación de Funciones de Densidad Beta*), respectivamente. Las probabilidades y cuantiles asociados con variables aleatorias de distribución gamma y beta se obtienen usando las aplicaciones *Gamma Probabilities and Quantiles* or *Beta Probabilities and Quantiles* (*Probabilidades y Cuantiles Gamma* o *Probabilidades y Cuantiles Beta).* Se proporcionan conjuntos de Ejercicios de Aplicaciones Breves para guiar al usuario a descubrir resultados interesantes e informativos asociados con variables aleatorias de distribución normal, beta y gamma (incluyendo las exponenciales y de χ^2). Mantenemos el énfasis en la distribución χ^2, incluyendo algunos resultados teóricos que son útiles en el posterior desarrollo de las distribuciones t y F.

En el Capítulo 5 se deja claro que las densidades condicionales no están definidas para valores de la variable condicionante donde la densidad marginal es cero. También hemos retenido el análisis de la "varianza condicional" y su uso para hallar la varianza de una variable aleatoria. Brevemente se estudian modelos jerárquicos. Al igual que en la edición anterior, el Capítulo 6 introduce el concepto del *apoyo* de una densidad y resalta que se puede usar un método de transformación cuando ésta es monótona en la región de apoyo. El método jacobiano está incluido para la implementación de una transformación bivariada.

En el Capítulo 7, la aplicación breve *Comparison of Student's t and Normal Distributions* (*Comparación de Distribuciones t de Student y Normales*) permite la visualización de similitudes y diferencias en funciones de densidad t y estándar normal y las aplicaciones breves *Probabilidades y Cuantiles Ji cuadrada, Probabilidades y Cuantiles t de Student* y las *Probabilidades y Cuantiles de Razón F* proporcionan probabilidades y cuantiles asociados con las distribuciones respectivas, todos con grados de libertad especificados por el usuario. La aplicación breve *DiceSample* utiliza el conocido ejemplo de lanzar dados a una mesa para introducir el concepto de una distribución de muestreo. Los resultados para diferentes tamaños muestrales permiten al usuario evaluar el impacto del tamaño de la muestra en la distribución muestral de la media. La aplicación breve también permite la visualización de la forma en que la distribución muestral es afectada si el dado no está balanceado. Bajo el encabezado general de "Distribuciones Muestrales y el Teorema del Límite Central", cuatro aplicaciones breves diferentes ilustran conceptos diferentes:

- *Básica* ilustra que, cuando se muestrea de una población normalmente distribuida, la media muestral está normalmente distribuida.
- *TamañoMuestral* exhibe el efecto del tamaño de la muestra en la distribución muestral de la media. La distribución muestral para dos tamaños muestrales (seleccionados por el usuario) son generados y mostrados simultáneamente uno al lado del otro. Las similitudes y diferencias de las distribuciones muestrales se hacen evidentes. Pueden generarse muestras de poblaciones con distribuciones "normales", uniformes, en forma de U y sesgadas. Las distribuciones asociadas que aproximan un muestreo normal pueden recubrirse en las distribuciones simuladas resultantes, permitiendo una inmediata evaluación visual de la calidad de la aproximación normal (vea Figura 3).
- *La Varianza* simula la distribución muestral de la varianza muestral cuando se muestrea de una población con una distribución "normal". La distribución teórica (proporcional a la de una variable aleatoria χ^2) puede recubrirse con el clic de un botón, dando de nuevo confirmación visual de que la teoría realmente funciona.
- *El Tamaño de Varianza* permite una comparación del efecto del tamaño muestral sobre la distribución de la varianza muestral (de nuevo, el muestreo de una población normal). La densidad teórica asociada puede recubrirse para ver que la teoría realmente funciona. Además, se ve que para grandes tamaños muestrales la varianza muestral tiene una distribución normal aproximada.

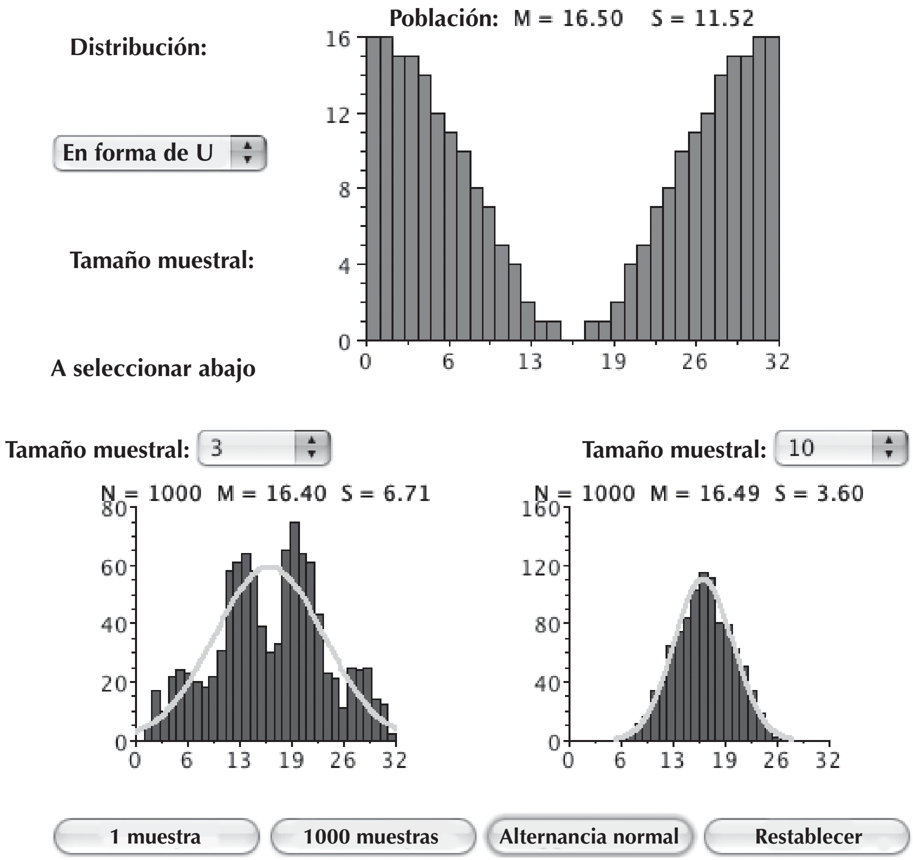


FIGURA **3** Ilustración de Aplicación Breve del teorema de límite central.

La aplicación *Normal Approximation to the Binomial* (*Aplicación Normal a la Binomial*) permite al usuario evaluar la calidad de la aproximación normal (continua) para probabilidades binomiales (discretas). Al igual que en capítulos anteriores, una secuencia de ejercicios de aplicación breve (applet) lleva al usuario a descubrir importantes e interesantes respuestas y conceptos. Desde un punto de vista más teórico, establecemos la independencia de la media muestral y varianza muestral para una muestra de tamaño 2 desde una distribución normal. Como antes, la prueba de este resultado para n general está contenida en un ejercicio opcional. Los ejercicios contienen deducciones paso a paso de la media y varianza para variables aleatorias con distribuciones t y F.

En todo el Capítulo 8 hemos resaltado las suposiciones asociadas con intervalos de confianza basados en las distribuciones t. También hemos incluido un breve examen de la robustez de los procedimientos t y la falta de éstos para los intervalos basados en las distribuciones χ^2 y F. La aplicación *ConfidenceIntervalP* (*IntervalodeConfianzaP*) ilustra propiedades de intervalos de confianza con muestras grandes para una proporción de población. En el Capítulo 9, las aplicaciones *PointSingle, PointbyPoint¸* y *PointEstimation* en última instancia llevan a una ilustración muy interesante de convergencia en probabilidad. En el Capítulo 10, la aplicación

Hypotesis Testing (for Proportions) ilustra conceptos importantes asociados con pruebas de hipótesis que incluyen lo siguiente:

- ¿Qué es lo que exactamente significa α?
- Pruebas basadas en tamaños de muestra más grande por lo general tienen probabilidades más pequeñas de cometer errores tipo II si el nivel de las pruebas permanece fijo.
- Para un tamaño muestral fijo, la función de potencia aumenta cuando el valor del parámetro se mueve más desde los valores especificados por la hipótesis nula.

Una vez que los usuarios visualicen estos conceptos, los subsiguientes desarrollos teóricos son más importantes y significativos. Las aplicaciones para las distribuciones χ^2, t, y F se usan para obtener valores p exactos para pruebas asociadas de hipótesis. También ilustramos de manera explícita que la potencia de una prueba uniformemente más poderosa puede ser menor (aun cuando es la más grande posible) de lo que se desea.

En el Capítulo 11, el modelo de regresión lineal simple se analiza en su totalidad (incluyendo intervalos de confianza, intervalos de predicción, y correlación) antes de introducir el método matricial al modelo de regresión lineal múltiple. Las aplicaciones *Fitting a Line Using Least Squares* y *Removing Points from Regression* ilustran que se logra el criterio de mínimos cuadrados y que unos cuantos puntos de información poco comunes pueden tener un considerable impacto en la línea de regresión ajustada. Los coeficientes de determinación y determinación múltiple se introducen, analizan y relacionan con las estadísticas t y F relevantes. Los ejercicios demuestran que los coeficientes altos (bajos) de valores de determinación (múltiples) no necesariamente corresponden a resultados que sean estadísticamente significativos (insignificantes).

El Capítulo 12 incluye una sección separada sobre el experimento de observaciones pareadas. Aunque se pueden usar numerosos conjuntos posibles de variables ficticias para llevar el análisis de varianza en un contexto de regresión, en el Capítulo 13 nos concentramos en las variables ficticias que por lo general se emplean en el SAS y otros paquetes de cómputo de análisis estadístico. El texto todavía se concentra básicamente en el diseño de bloques aleatorizado con efectos de bloque (no aleatorios). Si el profesor así lo desea, se puede usar una serie de ejercicios complementarios que se refieren al diseño de bloques aleatorizado con efectos de bloque aleatorio, para ilustrar las similitudes y diferencias de estas dos versiones del diseño de bloques aleatorizado.

El nuevo Capítulo 16 contiene una breve introducción a métodos de Bayes de inferencia estadística. El capítulo se concentra en el uso de datos y la distribución previa para obtener la posterior y en el uso de la posterior para obtener estimaciones, intervalos de credibilidad y pruebas de hipótesis para parámetros. La aplicación *Binomial Revision* (*Revisión Binomial*) facilita la comprensión del proceso por el que se usa información para actualizar la previa y obtener la posterior. Muchas de las distribuciones posteriores son distribuciones beta o gamma y las aplicaciones ya antes estudiadas son instrumentos para obtener intervalos de credibilidad o para calcular la probabilidad de varias hipótesis.

Los ejercicios

Esta edición contiene más de 350 nuevos ejercicios, muchos de los cuales usan las aplicaciones previamente citadas para guiar al usuario por una serie de pasos que llevan a más comprensión de conceptos importantes. Otros usan las aplicaciones breves (applets) para obtener intervalos de confianza o valores p que pudieran sólo aproximarse con el uso de tablas del apéndice. En la misma forma que en ediciones previas, parte de los nuevos ejercicios son teóricas mientras que otros contienen información de fuentes documentadas que se refieren a investigaciones realizadas en varios campos. Seguimos pensando que los ejercicios basados en datos reales, o en situaciones experimentales reales, permiten que el estudiante vea los usos prácticos de los diversos métodos estadísticos y de probabilística presentados en el texto. Al resolver estos ejercicios, los estudiantes adquieren conocimientos de aplicaciones reales de los resultados teóricos desarrollados en el texto. Este conocimiento hace que el aprendizaje de la teoría necesaria se disfrute más y produzca una más profunda comprensión de los métodos teóricos. Al igual que en ediciones previas, los ejercicios de grado de dificultad más alto están marcados con un asterisco (*). Las respuestas a los ejercicios de número impar aparecen al final del libro.

Tablas y apéndice

Hemos conservado el uso de tablas normales de cola superior porque los usuarios del texto las encuentran más cómodas. También hemos conservado el formato de la tabla de las distribuciones F que introdujimos en ediciones anteriores. Esta tabla de las distribuciones F dan valores críticos correspondientes a áreas de cola superior de .100, .050, .025, .010 y .005 en una sola tabla. Como las pruebas basadas en estadísticos que poseen la distribución F se presentan con mucha frecuencia, esta tabla facilita el cálculo de niveles de significancia alcanzados, o valores p, asociados con valores observados de estos estadísticos.

También hemos continuado nuestra práctica de dar fácil acceso a información que se usa con frecuencia. Debido a que las tablas t y normales son las tablas estadísticas que se usan con más frecuencia en el texto, en el Apéndice 3 del texto se dan copias de estas tablas. Los usuarios de ediciones anteriores con frecuencia nos han hablado favorablemente de la utilidad de tablas de las distribuciones comunes de probabilidad, medias, varianzas y funciones generadoras de momento dadas en el Apéndice 2 del libro. Además, en un suplemento al Apéndice 1 hemos incluido algunos resultados matemáticos que se usan con frecuencia. Estos resultados incluyen la expansión binomial de $(x + y)^n$, la expansión serie de e^x, sumas de series geométricas, definiciones de las funciones gamma y beta, y así sucesivamente. Al igual que antes, cada capítulo se inicia con un resumen que contiene los títulos de las principales secciones de ese capítulo.

Reconocimientos

Los autores desean agradecer a numerosos colegas, amigos y estudiantes que han dado útiles sugerencias respecto a los cambios de este texto. En particular, estamos en deuda con P. V. Rao, J. G. Saw, Malay Ghosh, Andrew Rosalsky y Brett Presnell por sus comentarios técnicos. Gary McClelland, Universidad de Colorado, hizo un excelente trabajo de desarrollo de las aplicaciones breves (applets) que se emplearon en el texto. Jason Owen, de la Universidad de Richmond, escribió el manual de soluciones; Mary Mortlock, Cal Poly, San Luis Obispo, comprobaron la exactitud de estas soluciones.

Deseamos agradecer a E. S. Pearson, W. H. Beyer, I. Olkin, R. A. Wilcox, C. W. Dunnett y A. Hald. Nos beneficiamos grandemente con las sugerencias de los revisores de las ediciones actual y previas del texto: Roger Abernathy, Arkansas State University; Elizabeth S. Allman, University of Southern Maine; Robert Berk, Rutgers University; Albert Bronstein, Purdue University; Subha Chakraborti, University of Alabama; Rita Chattopadhyay, Eastern Michigan University; EricChicken, Florida State University; Charles Dunn, Linfield College; Eric Eide, Brigham Young University; Nelson Fong, Creighton University; Dr. GailP. Greene, Indiana Wesleyan University; Barbara Hewitt, University of Texas, San Antonio; Richard Iltis, Willamette University; K. G. Janardan, Eastern Michigan University; Mark Janeba, Willamette University; Rick Jenison, Univeristy of Wisconsin, Madison; Jim Johnston, Concord University; Bessie H. Kirkwood, Sweet Briar College; Marc L. Komrosky, San Jose State University; Dr. Olga Korosteleva, California State University, Long Beach; Teck Ky, Evegreen Valley College; Matthew Lebo, Stony Brook University; Phillip Lestmann, Bryan College; Tamar London, Pennsylvania State University; Lisa Madsen, Oregon State University; Martin Magid, Wellesley College; Hosam M. Mahmoud, George Washington University; Kim Maier, Michigan State University; David W. Matolak, Ohio University; James Edward Mays, Virginia Commonwealth University; Katherine Mc Givney, Shippensburg Univesity; Sanjog Misra, Universityof Rochester; Donald F. Morrison, University of Pennsylvania, Wharton; Mir A. Mortazavi, Eastern New Mexico University; Abdel-Razzaq Mugdadi, Southern Illinois University; Ollie Nanyes, Bradley University; Joshua Naranjo, Western Michigan University; Sharon Navard, The College of New Jersey; Roger B. Nelsen, Lewis & Clark College; David K. Park, Washington University; Cheng Peng, University of Southern Maine; Selwyn Piramuthu, University of Florida, Gainesville; Robert Martin Price, Jr., East Tennessee State University; Daniel Rabinowitz, Columbia University; Julianne Rainbolt, Saint Louis University; Timothy A. Riggle, Baldwin-Wallace College; Mark Rizzardi, Humboldt State University; Jesse Rothstein, Princeton University; Katherine Schindler, Eastern Michigan University; Michael E. Schuckers, St. Lawrence University; Jean T. Sells, Sacred Heart University; Qin Shao, The University of Toledo; Alan Shuchat, Wellesley College; Laura J. Simon, Pennsylvania State University; Satyanand Singh, New York City College of Technology; Randall J. Swift, California State Polytechnic University, Pomona; David Sze, Monmouth University; Bruce E. Trumbo, California State University, East Bay; Harold Dean Victory, Jr., Texas Tech University; Thomas O. Vinson, Washington & Lee University; Vasant Waikar, Miami University, Ohio; Bette Warren, Eastern Michigan University; Steve White, Jacksonville State University; Shirley A. Wilson, North Central College; Lan Xue, OregonState University; y Elaine Zanutto, The Wharton School, University of Pennsylvania.

Deseamos también agradecer las contribuciones de Carolyn Crockett, nuestra editora; Catie Ronquillo, asistente de la editora; Ashley Summers, asistente editorial; Jennifer Liang, gerente de proyectos tecnológicos; Mandy Jellerichs, gerente de marketing; Ashley Pickering, asistente de marketing; y a todos los involucrados en la producción del texto; Hal Humphrey, gerente de producción; Betty Duncan, copyeditor, y Merrill Peterson y Sara Planck, coordinadoras de producción.

Finalmente, apreciamos el apoyo de nuestras familias durante el desarrollo de las varias ediciones de este texto.

DENNIS D. WACKERLY

WILLIAM MENDENHALL III

RICHARD L. SCHEAFFER

NOTA AL ESTUDIANTE

Como lo sugiere el título de *Estadística matemática con aplicaciones*, este libro se refiere a estadísticas, en teoría y aplicaciones, y sólo contiene matemáticas como una herramienta necesaria para dar al estudiante una firme comprensión de técnicas estadísticas. Las siguientes sugerencias para usar el texto aumentarán el aprendizaje y ahorrarán tiempo.

La conectividad del libro está dada por las introducciones y resúmenes de cada capítulo. Estas secciones explican la forma en que cada capítulo se ajusta en la imagen general de la inferencia estadística y cómo es que cada capítulo se relaciona con los precedentes.

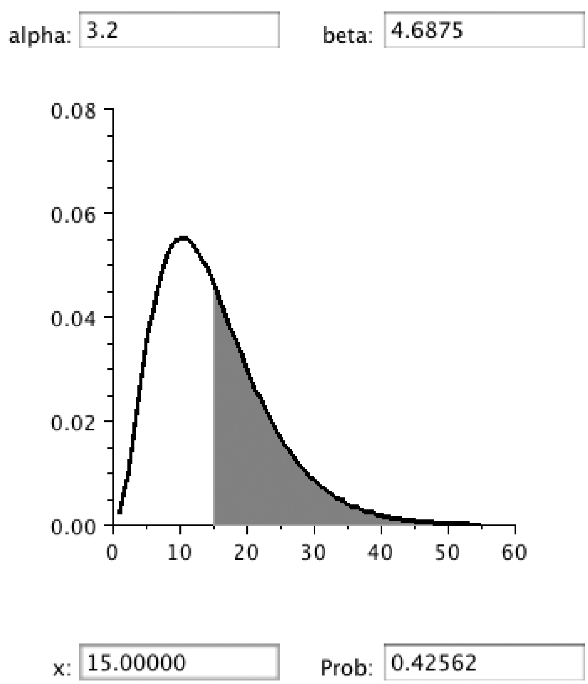


FIGURA **4** Cálculo de aplicación breve de la probabilidad de que una variable aleatoria de distribución gamma excede de su media

Dentro de los capítulos, se ponen de relieve importantes conceptos como definiciones. Éstas deben ser leídas varias veces hasta que sean entendidas con toda claridad porque forman la estructura sobre la que se construye todo lo demás. Los principales resultados teóricos se destacan como teoremas. Aun cuando no es necesario entender la prueba de cada uno de los teoremas, una clara comprensión del significado e implicaciones de éstos es esencial.

También es fundamental que el estudiante trabaje muchos de los ejercicios, al menos por cuatro razones:

- El lector puede estar seguro que entiende lo que ha leído sólo si pone su conocimiento a prueba en problemas de trabajo.
- Muchos de los ejercicios son de una naturaleza práctica y arrojan luz sobre las aplicaciones de probabilidad y estadística.
- Algunos de los ejercicios presentan nuevos conceptos y así amplían el material cubierto en el capítulo.
- Muchos de los ejercicios de aplicaciones breves ayudan a aumentar la intuición, facilitar la comprensión de conceptos y dar respuestas que no pueden (prácticamente) obtenerse con el uso de las tablas de los apéndices (vea la Figura 4).

D. D. W.

W. M.

R. L. S.

CAPÍTULO 1

¿Qué es estadística?

1.1 Introducción

Se emplean técnicas estadísticas en casi todas las fases de la vida. Se diseñan encuestas para recabar los primeros informes en un día de elecciones y pronosticar el resultado de una elección. Se hacen muestreos de consumidores para obtener información para predecir preferencias de productos. Médicos investigadores realizan experimentos para determinar el efecto de diversos medicamentos y condiciones ambientales controladas en seres humanos para inferir el tratamiento adecuado para varias enfermedades. Los ingenieros muestrean la característica de calidad de un producto y diversas variables de procesos controlables para identificar variables clave relacionadas con la calidad de un producto. Aparatos electrónicos recién manufacturados se muestrean antes de enviarlos para decidir si se embarcan o se mantienen lotes individuales. Los economistas observan varios índices del estado de la economía en un periodo y usan la información para pronosticar las condiciones de la economía en el futuro. Las técnicas estadísticas desempeñan un importante papel para alcanzar la meta de cada una de estas situaciones prácticas. El perfeccionamiento de la teoría que sirve de base a estas técnicas es el objetivo de este texto.

Un requisito previo para un examen de la teoría de estadística es una definición de *estadística* y un enunciado de sus objetivos. El *Webster's New Collegiate Dictionary* define estadística como "rama de las matemáticas que estudia la recolección, análisis, interpretación y presentación de masas de información numérica". Stuart y Ord (1991) expresan: "Estadística es la rama del método científico que estudia los datos obtenidos por contar o medir las propiedades de poblaciones". Rice (1995), comentando sobre experimentación y aplicaciones estadísticas, expresa que la estadística se "ocupa esencialmente de procedimientos para analizar información, en especial aquella que en algún sentido vago tenga un carácter aleatorio". Freund y Walpole (1987), entre otros, ven la estadística como una disciplina que abarca "la ciencia de basar inferencias en datos observados y todo el problema de tomar decisiones frente a una incertidumbre". Y Mood, Graybill y Boes 1974) definen la estadística como "la tecnología del método científico" y agregan que la estadística se ocupa de "(1) el diseño de experimentos e investigaciones, (2) inferencia estadística". Un examen superficial de estas definiciones sugiere una falta importante de acuerdo, pero todas poseen elementos comunes. Cada descripción implica que los datos se recolectan, con la inferencia como objetivo. Cada una requiere seleccionar un subconjunto de un gran conjunto de datos, ya sea existente o conceptual, para inferir las características del conjunto completo. Todos los autores implican que *la estadística es una teoría de información, siendo la inferencia su objetivo.*

La gran recopilación de datos que es el objetivo de nuestro interés se denomina *población*, y el subconjunto seleccionado de ella es una *muestra.* Las preferencias del electorado para un candidato gubernamental, Jones, expresadas en forma cuantitativa (1 por "prefieren" y 0 para "no prefieren") dan una población real, finita y existente de gran interés para Jones. Para determinar la verdadera fracción que está a favor de elegirlo, Jones necesitaría entrevistar a *todo* el electorado, trabajo que es prácticamente imposible. El voltaje en un punto particular en el sistema de guía para una nave espacial puede probarse en los únicos tres sistemas que se han construido. Los datos resultantes podrían usarse para estimar las características de voltaje para otros sistemas que podrían manufacturarse en el futuro. En este caso, la población es *conceptual.* Consideramos que la muestra de tres es representativa de una gran población de sistemas de guía que podrían construirse usando el mismo método. Presumiblemente, esta población tendría características similares a los tres sistemas de la muestra. Del mismo modo, las mediciones en los pacientes para un experimento médico representan una muestra de una población conceptual formada por todos los pacientes que de modo semejante sufren hoy, así como los que estarán afectados en el futuro próximo. El lector encontrará útil definir con claridad las poblaciones de interés para cada una de las situaciones descritas antes en esta sección y para aclarar la meta inferencial para cada una.

Es interesante observar que la industria y el gobierno de Estados Unidos gastan miles de millones de dólares cada año en busca de datos de experimentos, encuestas de muestreo y otros procedimientos de recolección de datos. Este dinero se gasta sólo para obtener información acerca de fenómenos susceptibles de medir en el ramo de finanzas, ciencias o las artes. Las implicaciones de esta afirmación dan la clave para la naturaleza de la muy valiosa aportación que la disciplina de estadística hace para la investigación y desarrollo en todos los campos de acción de la sociedad. La información útil para inferir alguna característica de una población (ya sea existente o conceptual) se compra en una cantidad especificada y resulta en una inferencia (estimación o decisión) con un *grado de bondad* asociado. Por ejemplo, si Jones hace arreglos para entrevistar una muestra de votantes, la información de la muestra se puede emplear para estimar la verdadera fracción de todos los votantes que están a favor de la elección de Jones. Además de la estimación, Jones debe estar interesado en la probabilidad (posibilidad) de que la estimación proporcionada sea cercana a la verdadera fracción de votantes elegibles que están a favor de su elección. De manera intuitiva, cuanto mayor sea el número de votantes elegibles de la muestra mayor será la probabilidad de una estimación precisa. Del mismo modo, si una decisión se toma con respecto a méritos relativos de dos procesos de manufactura basados en el examen de muestras de productos provenientes de ambos procesos, deberíamos estar interesados en la decisión respecto a cuál es mejor *y* en la probabilidad de que la decisión sea correcta. En general, el estudio de la estadística se ocupa del diseño de experimentos o encuestas muestrales para obtener una cantidad especificada de información al mínimo costo y del óptimo uso de esta información al hacer una inferencia acerca de una población. *La meta de la estadística es hacer una inferencia acerca de una población, con base en información contenida en una muestra de esa población y dar una medida de bondad asociada para la inferencia.*

Ejercicios

1.1 Para cada una de las siguientes situaciones, identifique la población de interés, la meta inferencial y diga cómo emprendería la recolección de una muestra.

a Un investigador universitario desea estimar la proporción de ciudadanos estadounidenses de la "generación X" que están interesados en iniciar sus propios negocios.

b Durante más de un siglo, la temperatura corporal normal en seres humanos ha sido aceptada como 37°C. ¿Es así realmente? Los investigadores desean estimar el promedio de temperatura de adultos sanos en Estados Unidos.

c Un ingeniero municipal desea estimar el promedio de consumo semanal de agua para unidades habitacionales unifamiliares en la ciudad.

d El National Highway Safety Council desea estimar la proporción de llantas para automóvil con dibujo o superficie de rodadura insegura, entre todas las llantas manufacturadas por una empresa específica durante el presente año de producción.

e Un politólogo desea estimar si la mayoría de los residentes adultos de un estado están a favor de una legislatura unicameral.

f Un científico del área médica desea estimar el tiempo promedio para que se vuelva a presentar cierta enfermedad.

g Un ingeniero electricista desea determinar si el promedio de vida útil de transistores de cierto tipo es mayor que 500 horas.

1.2 Caracterización de un conjunto de mediciones: métodos gráficos

En el sentido más amplio, hacer una inferencia implica describir en forma parcial o completa un fenómeno o un objeto físico. Se encuentra poca dificultad cuando se dispone de mediciones descriptivas apropiadas y significativas, pero éste no es siempre el caso. Por ejemplo, podríamos caracterizar una persona si usamos su estatura, peso, color de cabello y ojos, así como otras mediciones descriptivas de la fisonomía de la persona. Identificar un conjunto de mediciones descriptivas para caracterizar una pintura al óleo sobre tela sería un trabajo comparativamente más difícil. Caracterizar una población formada por un conjunto de mediciones es también muy difícil. En consecuencia, un preludio necesario para una discusión de hacer inferencias es la adquisición de un método para caracterizar un conjunto de números. Las caracterizaciones deben ser significativas para que el conocimiento de las mediciones descriptivas haga posible que visualicemos con claridad el conjunto de números. Además, requerimos que las caracterizaciones tengan significación práctica para que el conocimiento de las mediciones descriptivas de una población se pueda usar para resolver un problema práctico, no estadístico. Desarrollaremos nuestras ideas sobre este tema al examinar un proceso que genera una población.

Considere un estudio para determinar variables importantes que afectan las utilidades de una empresa que fabrica aparatos de maquinado especial o hechos bajo pedido. Algunas de estas variables podrían ser el importe del contrato en dólares, el tipo de industria con la cual se negocia el contrato, el grado de competencia para adquirir contratos, el vendedor que estima el contrato, costos fijos en dólares y el supervisor a quien se asigna el trabajo de organizar y dirigir el proceso de fabricación. El experto en estadística deseará medir la respuesta o variable dependiente, y la utilidad por contrato para varios trabajos (la muestra). Además de registrar la utilidad, el estadístico obtendrá mediciones sobre las variables que podrían estar relacionadas con la utilidad, las variables independientes. El objetivo de él o ella es usar información de la muestra para inferir la relación aproximada entre las variables independientes que acabamos de describir y la variable dependiente, la utilidad, y para medir la fuerza de esta relación. El objetivo del fabricante es determinar óptimas condiciones para maximizar las utilidades.

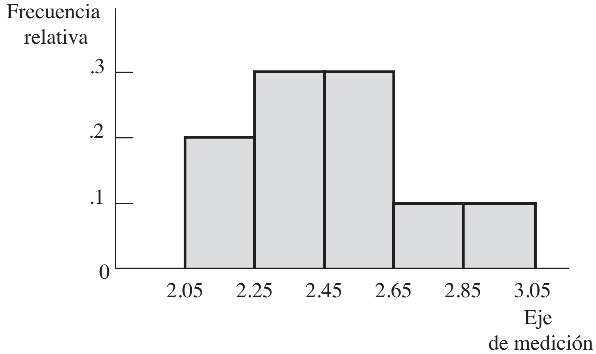
La población de interés del problema de manufactura es conceptual y está formado por todas las mediciones de utilidades (por unidad de capital y mano de obra invertidos) que podrían hacerse en contratos, ahora y en el futuro, para valores fijos de las variables independientes (tamaño del contrato, medida de competencia, etc.) Las mediciones de las utilidades van a variar de un contrato a otro en una forma aparentemente aleatoria como resultado de variaciones en materiales, tiempo necesario para completar segmentos individuales del trabajo y otras variables no controlables que afectan el trabajo. En consecuencia, vemos la población como la representada por una *distribución* de mediciones de utilidades, con la forma de la distribución dependiendo de valores específicos de las variables independientes. Nuestro deseo para determinar la relación entre la variable dependiente, la utilidad y un conjunto de variables independientes, por tanto, se convierte en un deseo para determinar el efecto de las variables independientes en la distribución conceptual de mediciones de población.

Una población individual (o cualquier conjunto de mediciones) puede estar caracterizada por una *distribución de frecuencia relativa*, que puede estar representada por un *histograma de frecuencia relativa*. Se construye una gráfica al subdividir el eje de medición en intervalos de igual ancho. Se construye un rectángulo sobre cada intervalo, de modo que la altura del rectángulo sea proporcional a la *fracción* del número total de mediciones que caen en cada celda. Por ejemplo, para caracterizar las diez mediciones 2.1, 2.4, 2.2, 2.3, 2.7, 2.5, 2.4, 2.6, 2.6 y 2.9, podríamos dividir el eje de medición en intervalos de igual ancho (por ejemplo .2 unidades), comenzando con 2.05. Las frecuencias relativas (fracción del número total de mediciones), calculadas para cada intervalo, se muestran en la Figura 1.1. Observe que la figura da una clara descripción gráfica de todo el conjunto de las diez mediciones.

Observe que no hemos dado reglas precisas para seleccionar el número, anchos o ubicaciones de los intervalos empleados para construir un histograma. Esto es porque la selección de estos elementos está un poco a discreción de la persona que intervenga en la construcción.

Aun cuando son arbitrarias, unas cuantas guías pueden ser muy útiles para seleccionar los intervalos. *Los puntos de subdivisión del eje de medición deben escogerse de modo que sea*

FIGURA **1.1** Histograma de frecuencia relativa



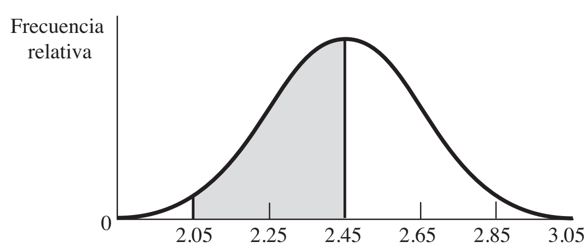
imposible que una medición caiga en un punto de división. Esto elimina una fuente de confusión y se obtiene con facilidad, como se indica en la Figura 1.1. La segunda guía se refiere al ancho de cada intervalo y, en consecuencia, al número mínimo de intervalos necesarios para describir los datos. Dicho en términos generales, deseamos obtener información sobre la forma de la distribución de los datos. Muchas veces la forma será como de montículo, como se ilustra en la Figura 1.2. (Otros prefieren referirse a estas distribuciones como forma de campana o normales.) El uso de numerosos intervalos con una pequeña cantidad de datos resulta en poco resumen y presenta una imagen muy semejante a los datos en su forma original. Cuanto mayor sea la cantidad de datos, el número de intervalos incluidos puede ser más grande pero al mismo tiempo se presenta una imagen satisfactoria de los datos. *Sugerimos abarcar el margen de los datos con* 5 *a* 20 *intervalos y usar el mayor número de intervalos para cantidades más grandes de datos.* En casi todas las aplicaciones prácticas se usan paquetes de software (Minitab, SAS, R, S+, JMP, etc.) para obtener cualesquiera histogramas deseados. Estos paquetes de cómputo producen histogramas que satisfacen restricciones ampliamente generalizadas sobre las que se han convenido escalas, número de intervalos empleados, anchos de intervalos y otros datos semejantes.

Algunos estadísticos piensan que la descripción de datos es un fin en sí misma. Con frecuencia se emplean histogramas para este propósito, pero hay muchos otros métodos gráficos que dan resúmenes significativos de la información contenida en un conjunto de datos. Algunas excelentes obras de consulta para el tema general de métodos descriptivos gráficos se dan en la bibliografía que aparece al final de este capítulo. Recuerde, no obstante, que el objetivo usual de la estadística es hacer inferencias. La distribución de frecuencia relativa asociada con un conjunto de datos y el histograma que la acompaña son suficientes para nuestros objetivos y para desarrollar el material en este texto. Esto se debe principalmente a la interpretación probabilística que se puede deducir del histograma de frecuencia, Figura 1.1. Ya hemos expresado que el área de un rectángulo para un intervalo dado es proporcional a la fracción del número total de mediciones que caen en dicho intervalo. Extendamos esta idea un paso más.

Si una medición se selecciona al azar del conjunto de datos original, la probabilidad de que caiga en un intervalo dado es proporcional al área bajo el histograma que se encuentre sobre ese intervalo. (En este punto nos apoyamos en el concepto de probabilidad del lego. Este término se examina con mayor detalle en el Capítulo 2.) Por ejemplo, para los datos empleados para construir la Figura 1.1, la probabilidad de que una medida seleccionada al azar caiga en el intervalo de 2.05 a 2.45 es .5 porque la mitad de las mediciones caen en este intervalo. De manera correspondiente, el *área bajo el histograma* de la Figura 1.1 sobre el intervalo de 2.05 a 2.45 es *la mitad del área total bajo el histograma.* Es claro que esta interpretación se aplica a la distribución de cualquier conjunto de mediciones, es decir, una población o una muestra.

Suponga que la Figura 1.2 da la distribución de frecuencia relativa de utilidades (en millones de dólares) para una población conceptual de respuestas de utilidades para contratos

FIGURA **1.2** Distribución de frecuencia relativa



en situaciones específicas de las variables independientes (tamaño de contrato, medida de competencia, etc.). La probabilidad de que el siguiente contrato (en las mismas situaciones de las variables independientes) dé una utilidad que caiga en el intervalo de 2.05 a 2.45 millones está dada por la proporción del área bajo la curva de distribución que está sombreada en la Figura 1.2.

Ejercicios

1.2 ¿Algunos poblados son más ventosos que otros? ¿Chicago merece el apodo de "la Ciudad de los Vientos"? A continuación aparece el promedio de velocidades del viento (en millas por hora) para 45 ciudades seleccionadas de Estados Unidos:

8.9	12.4	8.6	11.3	9.2	8.8	35.1	6.2	7.0
7.1	11.8	10.7	7.6	9.1	9.2	8.2	9.0	8.7
9.1	10.9	10.3	9.6	7.8	11.5	9.3	7.9	8.8
8.8	12.7	8.4	7.8	5.7	10.5	10.5	9.6	8.9
10.2	10.3	7.7	10.6	8.3	8.8	9.5	8.8	9.4

Fuente: The World Almanac and Book of Facts, 2004.

a Construya un histograma de frecuencia relativa para estos datos. (Seleccione las fronteras de clase sin incluir el valor de 35.1 en el intervalo de valores.)

b El valor 35.1 se registró en Mt. Washington, New Hampshire. ¿La geografía de esa ciudad explica la magnitud promedio de la velocidad de sus vientos?

c El promedio de velocidad de vientos en Chicago es 10.3 millas por hora. ¿Qué porcentaje de las ciudades tienen promedios de velocidad de vientos mayores que Chicago?

d ¿Piensa usted que Chicago tiene vientos inusuales?

1.3 De gran importancia para residentes de la región central de Florida es la cantidad de material radiactivo presente en el suelo de zonas recuperadas de la explotación minera de fosfatos. Las mediciones de la cantidad de ^{238}U en 25 muestras de suelo fueron como sigue (mediciones en picocurios por gramo):

.74	6.47	1.90	2.69	.75
.32	9.99	1.77	2.41	1.96
1.66	.70	2.42	.54	3.36
3.59	.37	1.09	8.32	4.06
4.55	.76	2.03	5.70	12.48

Construya un histograma de frecuencia relativa para estos datos.

1.4 Las 40 acciones principales del mercado secundario (OTC, por sus siglas en inglés), clasificadas por el porcentaje de acciones en circulación vendidas en un día el año pasado son como sigue:

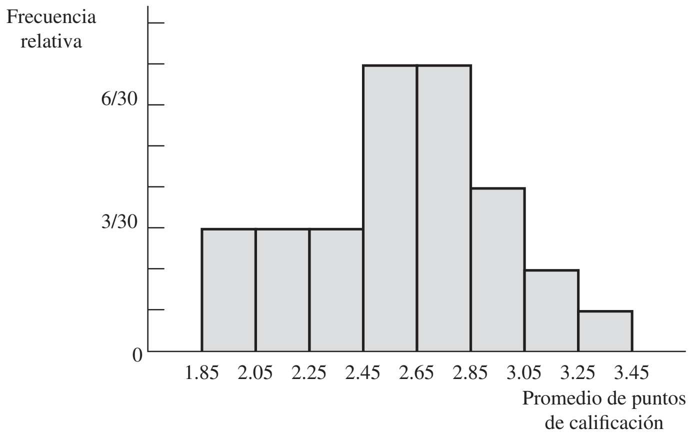
11.88	6.27	5.49	4.81	4.40	3.78	3.44	3.11	2.88	2.68
7.99	6.07	5.26	4.79	4.05	3.69	3.36	3.03	2.74	2.63
7.15	5.98	5.07	4.55	3.94	3.62	3.26	2.99	2.74	2.62
7.13	5.91	4.94	4.43	3.93	3.48	3.20	2.89	2.69	2.61

a Construya un histograma de frecuencia relativa para describir estos datos.

b ¿Qué proporción de estas 40 acciones principales vendió más de 4% de las acciones en circulación?

c Si una de las acciones se selecciona al azar de las 40 para las que se tomaron los datos precedentes, ¿cuál es la probabilidad de que venda menos de 5% de sus acciones en circulación?

1.5 Aquí damos el histograma de frecuencia relativa asociado con los promedios de puntos de calificación (PPC) de una muestra de 30 estudiantes:

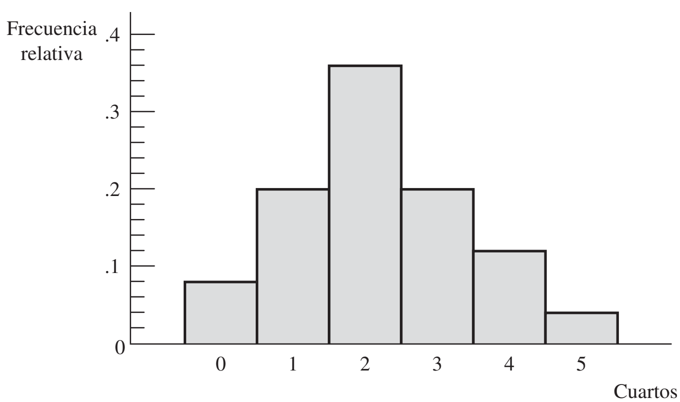


a ¿Cuáles de las categorías de PPC identificadas en el eje horizontal están asociadas con la proporción más grande de estudiantes?

b ¿Qué proporción de estudiantes tuvo PPC en cada una de las categorías que identificamos?

c ¿Qué proporción de los estudiantes tuvo PPC menores que 2.65?

1.6 El histograma de frecuencia relativa que aparece a continuación se construyó a partir de datos obtenidos de una muestra aleatoria de 25 familias. A cada una se le preguntó el número de cuartos de galón de leche que habían comprado la semana previa.

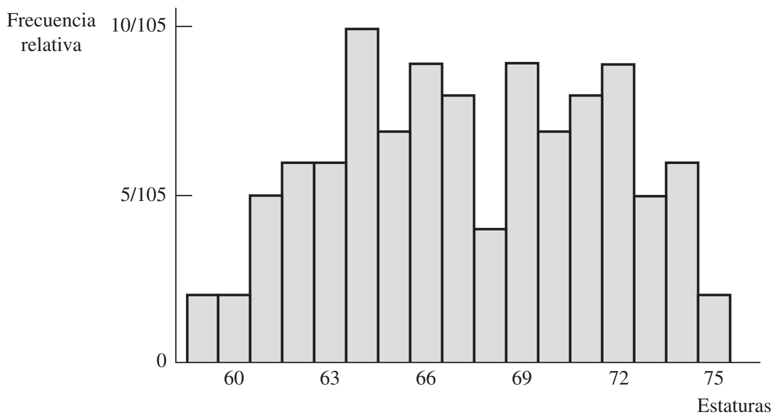


a Use este histograma de frecuencia relativa para determinar el número de cuartos de leche comprados por la proporción más grande de las 25 familias. La categoría asociada con la frecuencia relativa más grande se denomina *categoría modal*.

b ¿Qué proporción de las 25 familias compró más de 2 cuartos de leche?

c ¿Qué proporción compró más de 0 pero menos de 5 cuartos?

1.7 Las estaturas, declaradas por 105 estudiantes de un grupo de bioestadística se emplearon para construir el histograma que aparece a continuación.



a Describa la forma del histograma.

b ¿Este histograma tiene una característica poco común?

c ¿Puede el lector dar una explicación acerca de los dos picos del histograma? ¿Hay alguna consideración además de la estatura que resulta en los dos picos separados? ¿Cuál es?

1.8 Un artículo en *Archaeometry* presentó un análisis de 26 muestras de cerámica romano-británica hallada en cuatro hornos en sitios diferentes del Reino Unido. El porcentaje de óxido de aluminio en cada una de las 26 muestras aparece a continuación:

Llanederyn		Caldicot	Island Thorns	Ashley Rails
14.4	11.6	11.8	18.3	17.7
13.8	11.1	11.6	15.8	18.3
14.6	13.4		18.0	16.7
11.5	12.4		18.0	14.8
13.8	13.1		20.8	19.1
10.9	12.7			
10.1	12.5			

Fuente: A. Tubb, A. J. Parker y G. Nickless, "The Analysis of Romano-British Pottery by Atomic Absorption Spectrophotometry", en *Archaeometry* 22 (1980): 153.

a Construya un histograma de frecuencia relativa para describir el contenido de óxido de aluminio de las 26 muestras de cerámica.

b ¿Qué característica poco común observa el lector en este histograma? Al ver los datos, ¿puede dar una explicación de esta característica poco común?

1.3 Caracterización de un conjunto de mediciones: métodos numéricos

Los histogramas de frecuencia relativa presentados en la Sección 1.2 dan información útil respecto de la distribución de conjuntos de medición, pero los histogramas no suelen ser adecuados para el propósito de hacer inferencias. De hecho, numerosos histogramas similares podrían formarse a partir del mismo conjunto de mediciones. Para hacer inferencias acerca de una población basadas en información contenida en una muestra y para medir la bondad de las inferencias, necesitamos cantidades rigurosamente definidas para resumir la información contenida en una muestra. Estas cantidades muestrales por lo general tienen propiedades matemáticas, que se desarrollarán en capítulos siguientes, que nos permiten hacer enunciados de probabilidad con respecto a la bondad de nuestras inferencias.

Las cantidades que definimos son *medidas descriptivas numéricas* de un conjunto de datos. Buscamos algunos números que tienen interpretaciones significativas y que se pueden usar para describir la distribución de frecuencia de cualquier conjunto de mediciones. Centraremos nuestra atención en dos tipos de números descriptivos: *medidas de tendencia central* y *medidas de dispersión o variación.*

Es probable que la medida de tendencia central más común empleada en estadística sea la media aritmética. (Debido a que éste es el único tipo de media que estudiamos en este texto, omitiremos la palabra *aritmética.*)

DEFINICIÓN **1.1**

La *media* de una muestra de n respuestas medidas $y_1, y_2, \ldots, y_n$ está dada por

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

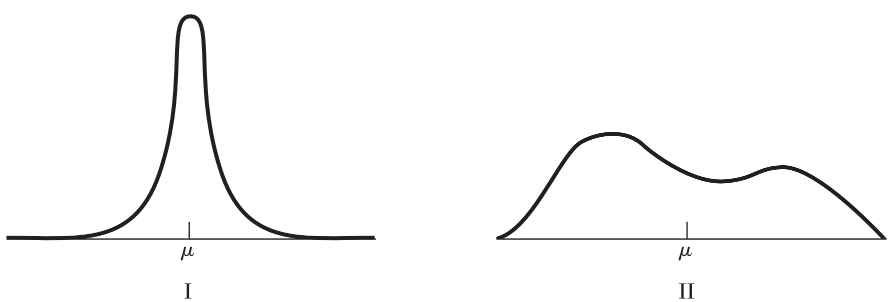
La media poblacional correspondiente se denota como μ.

El símbolo $\overline{y}$, que se lee "*y* barra", se refiere a una media muestral. Por lo general no podemos medir el valor de la media poblacional, μ; más bien, μ es una constante desconocida que podemos estimar usando información muestral.

La media de un conjunto de mediciones sólo localiza el centro de la distribución de datos, por sí misma no proporciona una descripción adecuada de un conjunto de mediciones. Dos conjuntos de mediciones podrían tener distribuciones de frecuencia muy diferentes pero iguales medias, como se ve en la Figura 1.3. La diferencia entre las distribuciones I y II de la figura se encuentra en la variación o dispersión de las mediciones que están a lado y lado de la media. Para describir en forma adecuada los datos, también debemos definir medidas de la variabilidad de datos.

La medida de variabilidad más común empleada en estadística es la varianza, que es una función de las desviaciones (o distancias) de las mediciones muestrales desde la media.

FIGURA **1.3** Distribuciones de frecuencia con iguales medias pero con diferentes cantidades de variación



DEFINICIÓN 1.2

La *varianza* de una muestra de mediciones $y_1, y_2, \ldots, y_n$ es la suma del cuadrado de las diferencias entre las mediciones y su media, dividida entre $n - 1$. Simbólicamente, la varianza muestral es

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

La correspondiente varianza poblacional está denotada por el símbolo σ^2.

Observe que dividimos entre $n - 1$ en lugar de entre n en nuestra definición de s^2. La razón teórica para esta elección de divisor se encuentra en el Capítulo 8, donde demostraremos que s^2 definida en esta forma proporciona un "mejor" estimador para la verdadera varianza de la población, σ^2. No obstante, es útil considerar a s^2 como "casi" el promedio del cuadrado de las desviaciones de los valores observados desde su media. Cuanto mayor sea la varianza de un conjunto de mediciones, mayor será la cantidad de variación dentro del conjunto. La varianza es de valor al comparar la variación relativa de dos conjuntos de mediciones, pero proporciona información acerca de la variación en un solo conjunto únicamente cuando se interpreta en términos de la desviación estándar.

DEFINICIÓN 1.3

La *desviación estándar* de una muestra de mediciones es la raíz cuadrada positiva de la varianza; esto es,

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

La correspondiente desviación estándar *poblacional* está denotada por $\sigma = \sqrt{\sigma^2}$.

Aun cuando está estrechamente relacionada con la varianza, la desviación estándar se puede usar para dar una imagen más o menos precisa de la variación de datos para un solo conjunto de mediciones. Se puede interpretar usando el teorema de Tchebysheff (que se estudia en el Ejercicio 1.32 y se presentará formalmente en el Capítulo 3) y por la regla empírica (que ahora explicamos).

Muchas distribuciones de datos de la vida real tienen forma de montículo; esto es, se pueden aproximar por medio de una distribución de frecuencia en forma de campana conocida como curva normal. Los datos que poseen distribuciones en forma de montículo tienen características definidas de variación, como se expresa en el enunciado siguiente.

Regla empírica

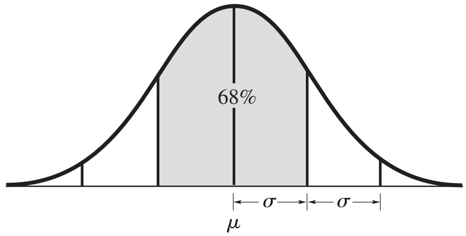
Para una distribución de mediciones que sea aproximadamente normal (forma de campana), se deduce que el intervalo con puntos extremos

$\mu \pm \sigma$ contiene aproximadamente 68% de las mediciones.

$\mu \pm 2\sigma$ contiene aproximadamente 95% de las mediciones.

$\mu \pm 3\sigma$ contiene casi todas las mediciones.

FIGURA 1.4
Curva normal



Como se mencionó en la Sección 1.2, una vez que se conozca la distribución de frecuencia de un conjunto de mediciones, se pueden hacer enunciados de probabilidad respecto de las mediciones. Estas probabilidades se mostraron como áreas bajo un histograma de frecuencia. En forma análoga, las probabilidades especificadas en la regla empírica son áreas bajo la curva normal mostrada en la Figura 1.4.

El uso de la regla empírica se ilustra mediante el siguiente ejemplo. Suponga que se sabe que las calificaciones en un examen vocacional aplicado a todos los estudiantes de último año de preparatoria en un estado tienen, aproximadamente, una distribución normal con media $\mu = 64$ y desviación estándar $\sigma = 10$. Entonces se puede deducir que aproximadamente 68% de las calificaciones están entre 54 y 74, que aproximadamente 95% de las calificaciones están entre 44 y 84 y que casi todas las calificaciones están entre 34 y 94. Así, el conocimiento de la media y la desviación estándar nos da una imagen más o menos buena de la distribución de frecuencia de las calificaciones.

Supongamos que se elige al azar un solo estudiante de preparatoria entre los que hicieron el examen. ¿Cuál es la probabilidad de que su resultado esté entre 54 y 74? Basándonos en la regla empírica, encontramos que 0.68 es una respuesta razonable a esta pregunta de probabilidad

La utilidad y valor de la regla empírica se deben a la ocurrencia común de distribuciones aproximadamente normales de datos en la naturaleza, aún más porque la regla se aplica a distribuciones que no son exactamente normales pero sí en forma de montículo. El estudiante encontrará que aproximadamente 95% de un conjunto de mediciones estará dentro de 2σ de μ para una variedad de distribuciones.

Ejercicios

1.9 Los ritmos de respiración en reposo para estudiantes en edad universitaria están normalmente distribuidos, en forma aproximada, con una media de 12 y desviación estándar de 2.3 respiraciones por minuto. ¿Qué fracción de todos los estudiantes en edad universitaria tienen ritmos de respiración en los intervalos siguientes?

a 9.7 a 14.3 respiraciones por minuto

b 7.4 a 16.6 respiraciones por minuto

c 9.7 a 16.6 respiraciones por minuto

d Menos de 5.1 o más de 18.9 respiraciones por minuto

1.10 Se ha proyectado que el promedio y la desviación estándar del tiempo empleado en línea usando la Internet son, respectivamente, 14 y 17 horas por persona por año (muchos no usan la Internet en absoluto).

a ¿Qué valor está exactamente a 1 desviación estándar debajo de la media?

b Si el tiempo empleado en línea usando la Internet está normalmente distribuido en forma aproximada, ¿qué proporción de los usuarios pasa un tiempo en línea que es menor al valor hallado en el inciso a?

c ¿El tiempo empleado en línea, usando la Internet, está normalmente distribuido en forma aproximada? ¿Por qué?

1.11 Los siguientes resultados en sumatorias nos ayudarán a calcular la varianza muestral s^2. Para cualquier constante c,

a $\sum_{i=1}^{n} c = nc.$

b $\sum_{i=1}^{n} cy_i = c \sum_{i=1}^{n} y_i.$

c $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i.$

Use a, b y c para demostrar que

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i \right)^2 \right].$$

1.12 Use el resultado del Ejercicio 1.11 para calcular s para las $n = 6$ mediciones muestrales 1, 4, 2, 1, 3 y 3.

1.13 Consulte el Ejercicio 1.2.

a Calcule $\overline{y}$ y s para los datos dados.

b Calcule el intervalo $\overline{y} \pm ks$, para $k = 1$, 2 y 3. Cuente el número de mediciones que caen dentro de cada intervalo y compare este resultado con el número que usted esperaría de acuerdo con la regla empírica.

1.14 Consulte el Ejercicio 1.3 y repita los incisos a y b del Ejercicio 1.13.

1.15 Consulte el Ejercicio 1.4 y repita los incisos a y b del Ejercicio 1.13.

1.16 En el Ejercicio 1.4 hay un valor extremadamente grande (11.88). Elimine este valor y calcule $\overline{y}$ y s para las restantes 39 observaciones. Del mismo modo, calcule los intervalos $\overline{y} \pm ks$ para $k = 1$, 2 y 3; cuente el número de mediciones en cada uno; luego compare estos resultados con los pronosticados por la regla empírica. Compare las respuestas de aquí con las halladas en el Ejercicio 1.15. Considere el efecto de una sola observación grande sobre $\overline{y}$ y s.

1.17 El *rango* de un conjunto de mediciones es la diferencia entre los valores máximo y mínimo. La regla empírica sugiere que la desviación estándar de un conjunto de mediciones puede ser casi aproximada en un cuarto de la amplitud (esto es, amplitud/4). Calcule esta aproximación a s para los conjuntos de datos de los Ejercicios 1.2, 1.3 y 1.4. Compare el resultado en cada caso con el valor calculado real de s.

1.18 Los exámenes de aptitud escolar verbales y matemáticos del Consejo Universitario se califican en una escala de 200 a 800. Parece razonable suponer que la distribución de calificaciones es aproximadamente normal para ambos exámenes. Use el resultado del Ejercicio 1.17 para calcular la desviación estándar para las calificaciones del examen verbal.

1.19 De acuerdo con la Environmental Protection Agency, el cloroformo, que en su forma gaseosa es un agente cancerígeno, está presente en pequeñas cantidades en las 240,000 fuentes públicas de agua de Estados Unidos. Si la media y la desviación estándar de las cantidades de cloroformo presentes en las fuentes de agua son 34 y 53 microgramos por litro (μg/L), respectivamente, explique por qué las cantidades de cloroformo no tienen una distribución normal.

1.20 Los costos semanales de mantenimiento para una fábrica, registrados en un largo periodo y ajustados a la inflación, tienden a tener una distribución aproximadamente normal con un promedio de \$420 y una desviación estándar de \$30. Si la cantidad de \$450 está presupuestada para la semana próxima, ¿cuál es una probabilidad aproximada de que se rebase esta cantidad presupuestada?

1.21 El fabricante de un nuevo aditivo de alimento para ganado dice que 80% de los animales alimentados con una dieta que incluya este aditivo deben tener un aumento mensual de peso de más de 20 libras. Una muestra grande de mediciones de aumento de peso para ganado alimentado con esta dieta muestra una distribución aproximadamente normal, con media de 22 libras y desviación estándar de 2 libras. ¿Piensa el lector que la información muestral contradice lo dicho por el fabricante? (Calcule la probabilidad de un aumento de peso mayor que 20 libras.)

1.4 Forma en que se hacen inferencias

El mecanismo para hacer inferencias se puede ilustrar bien al analizar nuestros propios procedimientos intuitivos para hacer inferencias.

Suponga que dos candidatos compiten para un cargo público en nuestra comunidad y que deseamos determinar si nuestro candidato, Jones, es preferido para ganar. La población de interés es el conjunto de respuestas de todos los votantes elegibles que votarán el día de la elección y deseamos determinar si la fracción que apoya a Jones excede de .5. Para más simplicidad, suponga que todos los votantes elegibles irán a las urnas y que al azar seleccionamos una muestra de 20 de la lista de votantes. Todos ellos fueron entrevistados y todos están a favor de Jones. ¿Qué concluimos acerca de las posibilidades de Jones para ganar la elección?

Hay poca duda de que casi todos nosotros inferimos de inmediato que Jones ganará. Ésta es una inferencia fácil de hacer, pero no es nuestra meta inmediata. Más bien, deseamos examinar los procesos mentales que se emplearon para llegar a esta conclusión acerca del comportamiento esperado de una gran población de votantes basada en una muestra de sólo 20 personas.

Ganar significa obtener más de 50% de los votos. ¿Concluimos que Jones ganaría porque pensamos que la fracción de la muestra que apoya a Jones era idéntica a la fracción que apoya a Jones en la población? Sabemos que probablemente esto no sea cierto. Un experimento sencillo verificará que la fracción de la muestra que apoya a Jones no necesita ser igual que la fracción de la población que lo apoya. Si se lanza al aire una moneda, es intuitivamente obvio que la verdadera proporción de veces que caiga con la cara hacia arriba es .5, pero si muestreamos los resultados de nuestra moneda lanzándola al aire 20 veces, la proporción de caras variará de una muestra a otra; esto es, en una ocasión podríamos observar 12 caras de 20 tiros, para una proporción de $12/20 = .6$. En otra ocasión, podríamos observar 8 caras en 20 tiros, para una proporción muestral de $8/20 = .4$. De hecho, la proporción muestral de caras podría ser $0, .05, .10, \ldots, 1.0$.

¿Concluimos que Jones ganaría porque sería imposible que 20 de cada 20 votantes muestrales lo apoyaran si de hecho menos de 50% del electorado trató de votar por él? La respuesta a esta pregunta es un no rotundo, pero da la clave para nuestra línea oculta de lógica. No es *imposible* sacar 20 de cada 20 que apoyan a Jones cuando menos de 50% del electorado lo apoya, pero es *altamente improbable.* Como resultado de ello, concluimos que ganaría.

Este ejemplo ilustra el importante papel que desempeña la probabilidad al hacer inferencias. Los probabilistas suponen que conocen la estructura de la población de interés y usan la teoría de probabilidad para calcular la probabilidad de obtener una muestra particular. Suponiendo que conocen la estructura de una población generada al sacar al azar cinco cartas de una baraja, los probabilistas calculan la probabilidad de que el saque dará tres ases y dos reyes. Algunos estadísticos usan la probabilidad para hacer el viaje a la inversa, es decir, de la muestra a la población. Al observar cinco ases en una muestra de cinco cartas, de inmediato infieren que la baraja (que genera la población) está "cargada" y no es estándar. La probabilidad de sacar cinco ases de una baraja estándar es cero, que es un caso exagerado pero comprueba la idea. Para hacer inferencias es básico el problema de calcular la probabilidad de una muestra observada y, en consecuencia, la probabilidad es un mecanismo que se emplea para hacer inferencias estadísticas.

Es oportuno un comentario final. Si usted no pensó que la muestra justificaba la inferencia de que Jones ganaría, no se sienta contrariado. Podemos confundirnos fácilmente al hacer evaluaciones intuitivas de las probabilidades de eventos. Si usted decidió que la probabilidad era muy baja de que 20 votantes de entre 20 estuvieran a favor de Jones, suponiendo que éste perdiera, tenía razón, pero no es difícil elaborar un ejemplo en el que una evaluación intuitiva de probabilidad sea errónea. Las evaluaciones intuitivas de probabilidades no son satisfactorias y necesitamos una rigurosa teoría de probabilidad para desarrollar métodos de inferencia.

1.5 Teoría y realidad

Las teorías son conjeturas propuestas para explicar fenómenos del mundo real. Como tales, las teorías son aproximaciones o modelos de la realidad. Estos modelos o explicaciones de la realidad se presentan en forma verbal en algunos campos menos cuantitativos y como relaciones matemáticas en otros. Mientras que una teoría de cambio social podría expresarse verbalmente en sociología, una descripción del movimiento de una cuerda en vibración se presenta de un modo matemático preciso en física. Cuando escogemos un modelo matemático para un proceso físico, esperamos que el modelo refleje fielmente, en términos matemáticos, los atributos del proceso físico. Si es así, el modelo matemático se puede usar para llegar a conclusiones acerca del proceso mismo. Si pudiéramos desarrollar una ecuación para predecir la posición de una cuerda en vibración, la calidad de la predicción dependería de lo bien que se ajuste la ecuación al movimiento de la cuerda. El proceso de hallar una buena ecuación no es necesariamente sencillo y por lo general requiere de varias suposiciones de simplificación (masa uniforme de la cuerda, ausencia de resistencia del aire, etc.). El criterio final para decidir si un modelo es "bueno" es si da información buena y útil. La motivación para usar modelos matemáticos radica esencialmente en su utilidad.

Este texto se ocupa de la teoría de estadística y por tanto tiene que ver con modelos de la realidad. Postularemos distribuciones de frecuencia teóricas para poblaciones y desarrollaremos una teoría de probabilidad e inferencia de un modo matemático preciso. El resultado neto será un modelo teórico o matemático para adquirir y utilizar información de la vida real. El modelo no será una representación exacta en la naturaleza, pero esto no debe molestarnos. Su utilidad, como la de otras teorías, se medirá por su capacidad para ayudarnos a entender la naturaleza y a resolver problemas del mundo real.

1.6 Resumen

El objetivo de la estadística es hacer una inferencia acerca de una población, con base en información contenida en una muestra tomada de esa población. La teoría de la estadística es una teoría de información que se ocupa de cuantificar información, diseñar experimentos o procedimientos para recolectar datos y analizarlos. Nuestra meta es minimizar el costo de una cantidad específica de información y usar esta información para hacer inferencias. Lo que es más importante es que hemos visto hacer una inferencia acerca de la población desconocida como un procedimiento de dos pasos. Primero, hacemos una lista de un procedimiento inferencial apropiado para la situación dada. En segundo término buscamos una medida de la bondad de la inferencia resultante. Por ejemplo, toda estimación de una característica de población basada en información contenida en la muestra podría tener asociado con ella un límite probabilístico en el error de estimación.

Un preludio necesario para hacer inferencias acerca de una población es la capacidad para describir un conjunto de números. Las distribuciones de frecuencia dan un método gráfico y útil para caracterizar poblaciones conceptuales o reales de números. Las medidas descriptivas numéricas son a veces más útiles cuando deseamos hacer una inferencia y medir la bondad de esa inferencia.

El mecanismo para hacer inferencias es proporcionado por la teoría de probabilidad. El probabilista razona desde una población conocida hasta el resultado de un experimento individual, la muestra. En contraste, el experto en estadística utiliza la teoría de probabilidad para calcular la probabilidad de una muestra observada y para inferir a partir de ésta las características de una población desconocida. Así, la probabilidad es la base de la teoría de estadística.

Por último, hemos observado la diferencia entre teoría y realidad. En este texto estudiaremos la teoría matemática de la estadística, que es una idealización de la naturaleza. Es rigurosa, matemática y sujeta a estudio en un vacío aislado por completo del mundo real. O puede estar ligada estrechamente a la realidad y puede ser útil para hacer inferencias a partir de datos en todos los campos de la ciencia. En este texto seremos utilitarios. No veremos la estadística como una rama de las matemáticas sino como un campo de las ciencias que se ocupa en desarrollar una teoría práctica de información. Consideraremos la estadística como un campo separado, análogo a la física, no como rama de las matemáticas sino como una teoría de la información que utiliza gran cantidad de matemáticas.

Los capítulos siguientes profundizarán en los temas que hemos encontrado en esta introducción. Empezaremos con un estudio del mecanismo empleado para hacer inferencias, la teoría de probabilidad. Esta teoría proporciona modelos teóricos para generar datos experimentales y, por tanto, da la base para nuestro estudio de inferencia estadística.

Bibliografía y lecturas adicionales

Cleveland, W. S.1994. *The Elements of Graphing Data*. Murray Hill, N. J.: AT&T Bell Laboratories.

———. *Visualizing Data*. 1993. Summit, N. J.: Hobart Press.

Fraser, D. A. S. 1958. *Statistics, an Introduction*. New York: Wiley.

Freund, J. E., and R. E. Walpole. 1987. *Mathematical Statistics*, 4th ed. Englewood Cliffs, N. J.: Prentice Hall.

Iman, R. L. 1994. A *Data-Based Approach to Statistics*. Belmont, Calif.: Duxbury Press.

Mendenhall, W., R. J. Beaver, and B. M. Beaver. 2006. *Introduction to Probability and Statistics*, 12th ed. Belmont, Calif.: Duxbury Press.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill.

Moore, D. S., and G. P. McCabe. 2002. *Introduction to the Practice of Statistics*, 4th ed. New York: Freeman.

Rice, J. A. *Mathematical Statistics and Data Analysis*, 2nd ed. Belmont, Calif.: Duxbury Press, 1995.

Stuart, A., and J. K. Ord. 1991. *Kendall's Theory of Statistics*, 5th ed., vol. 1. London: Edward Arnold.

Ejercicios complementarios

1.22 Demuestre que la suma de las desviaciones de un conjunto de mediciones alrededor de su media es igual a cero; es decir,

$$\sum_{i=1}^{n}(y_i - \overline{y}) = 0.$$

1.23 La duración media de anuncios comerciales por televisión es de 75 segundos con desviación estándar de 20 segundos. Para contestar lo siguiente, suponga que las duraciones están distribuidas normalmente en forma aproximada.

a ¿Qué porcentaje de anuncios dura más de 95 segundos?

b ¿Qué porcentaje de anuncios dura entre 35 y 115 segundos?

c ¿Esperaría que los anuncios duren más de 2 minutos? ¿Por qué sí o por qué no?

1.24 Los deportes acuáticos se han sugerido como método de ejercicio cardiovascular para atletas lesionados y otros que deseen un programa de entrenamiento aeróbico de bajo impacto. En un estudio para investigar la relación entre cadencia de ejercicios y ritmo cardiaco,[1] se midieron los ritmos cardiacos de 20 voluntarios en buenas condiciones a una cadencia de 48 ciclos por minuto (un ciclo formado por dos pasos). Los datos son como sigue:

87	109	79	80	96	95	90	92	96	98
101	91	78	112	94	98	94	107	81	96

a Use la amplitud de las mediciones para obtener una estimación de la desviación estándar.

b Construya un histograma de frecuencia para los datos. Use el histograma para obtener una aproximación visual a $\overline{y}$ y s.

c Calcule $\overline{y}$ y s. Compare estos resultados con las verificaciones de cálculo proporcionadas por los incisos a y b.

d Construya los intervalos $\overline{y} \pm ks$, $k = 1, 2$ y 3, y cuente el número de mediciones que caen en cada intervalo. Compare las fracciones que caen en los intervalos con las fracciones que esperaría de acuerdo con la regla empírica.

1. R. P. Wilder, D. Breenan y D. E. Schotte, "A Standard Measure for Exercise Prescription for Aqua Running", en *American Journal of Sports Medicine* 21(1) (1993): 45.

1.25 Los datos siguientes dan los tiempos de falla para $n = 88$ radiotransmisores-receptores:

16	224	16	80	96	536	400	80
392	576	128	56	656	224	40	32
358	384	256	246	328	464	448	716
304	16	72	8	80	72	56	608
108	194	136	224	80	16	424	264
156	216	168	184	552	72	184	240
438	120	308	32	272	152	328	480
60	208	340	104	72	168	40	152
360	232	40	112	112	288	168	352
56	72	64	40	184	264	96	224
168	168	114	280	152	208	160	176

a Utilice el rango para calcular s para los $n = 88$ tiempos para falla.

b Construya un histograma de frecuencia para los datos. [Observe la tendencia de la distribución a extenderse hacia fuera (sesgo) a la derecha.]

c Use una calculadora (o computadora) para calcular $\overline{y}$ y s. (Hacer manualmente el cálculo es demasiado tedioso para este ejercicio.)

d Calcule los intervalos $\overline{y} \pm ks$, $k = 1$, 2 y 3, y cuente el número de mediciones que caen en cada intervalo. Compare sus resultados con los resultados de la regla empírica. Observe que la regla empírica proporciona una descripción más o menos buena de estos datos, aun cuando la distribución está altamente sesgada.

1.26 Compare la razón entre la amplitud y s para los tres tamaños muestrales ($n = 6$, 20 y 88) para los Ejercicios 1.12, 1.24 y 1.25. Observe que la razón tiende a aumentar a medida que aumenta la cantidad de datos. Cuanto mayor es la cantidad de datos, mayor será su tendencia a contener algunos valores extremos que inflarán el rango y tendrán relativamente poco efecto en s. Pasamos por alto este fenómeno y sugerimos que el lector use 4 como la razón para hallar un valor estimado de s al comprobar sus cálculos.

1.27 Un conjunto de 340 calificaciones de examen, que muestran una distribución de frecuencia relativa en forma de campana, tiene una media de $\overline{y} = 72$ y una desviación estándar de $s = 8$. ¿Aproximadamente cuántas de las calificaciones se esperaría que cayeran en el intervalo de 64 a 80? ¿Y en el intervalo de 56 a 88?

1.28 La descarga de sólidos suspendidos de una mina de fosfato está normalmente distribuida con una descarga media diaria de 27 miligramos por litro (mg/L) y desviación estándar de 14 mg/L. ¿En qué proporción de los días será la descarga diaria menor que 13 mg/L?

1.29 Una máquina produce cojinetes con diámetro medio de 3.00 pulgadas y desviación estándar de 0.01 pulgada. Los cojinetes con diámetros de más de 3.02 o de menos de 2.98 no satisfacen las especificaciones de calidad.

a ¿Aproximadamente qué fracción de la producción de esta máquina no cumplirá con las especificaciones?

b ¿Qué suposiciones hizo usted con respecto a la distribución de diámetros de cojinetes para contestar esta pregunta?

1.30 En comparación con sus compañeras que se quedan en casa, las mujeres empleadas fuera de casa tienen niveles más altos de lipoproteínas de alta densidad (HDL, por sus siglas en inglés), el colesterol "bueno" asociado con menor riesgo de ataques al corazón. Un estudio de niveles de colesterol en 2000 mujeres, en edades de 25 a 64 y que viven en Augsburg, Alemania, fue realizado por Ursula Haertel, Ulrigh Keil

2. *Science News* 135 (junio de 1989): 389.

y colegas[2] en el GSF-Medis Institut en Munich. De estas 2000 mujeres, el 48% que trabajaban fuera de casa tuvieron niveles de HDL que estaban entre 2.5 y 3.6 miligramos por decilitro (mg/dL) más altos que los niveles de HDL de sus similares que se quedaban en casa. Suponga que la diferencia en niveles de HDL está normalmente distribuida, con media 0 (lo que indica que no hay diferencia entre los dos grupos de mujeres) y desviación estándar de 1.2 mg/dL. Si usted fuera a seleccionar al azar una mujer empleada y una que se queda en casa, ¿cuál es la probabilidad de que la diferencia en sus niveles de HDL esté entre 1.2 y 2.4?

1.31 El año pasado un proceso de producción de fertilizantes ha mostrado un promedio de producción diaria de 60 toneladas con una varianza en la producción diaria de 100. Si la producción bajara a menos de 40 toneladas mañana, ¿este resultado le haría sospechar una anomalía en el proceso? (Calcule la probabilidad de obtener menos de 40 toneladas.) ¿Qué suposiciones hizo usted con respecto a la distribución de producciones?

***1.32** Sea $k \geq 1$. Demuestre que, para cualquier conjunto de n mediciones, la fracción incluida en el intervalo $\overline{y} - ks$ a $\overline{y} + ks$ es al menos $(1 - 1/k^2)$. [*Sugerencia:*

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right].$$

En esta expresión, sustituya con ks todas las desviaciones para las cuales $|y_i - \overline{y}| \geq ks$. Simplifique.] Este resultado se conoce como *teorema de Tchebysheff*.[3]

1.33 El gerente de personal de cierta empresa tiene registros del número de empleados ausentes por día. El número promedio de ausentes es 5.5 y la desviación estándar es 2.5. Debido a que hay muchos días con cero, uno o dos ausentes y sólo unos pocos con más de diez ausentes, la distribución de frecuencia está altamente sesgada. El gerente desea publicar un intervalo en el que al menos 75% de estos valores mienta. Use el resultado del Ejercicio 1.32 para hallar ese intervalo.

1.34 Para los datos estudiados en el Ejercicio 1.33, dé un límite superior a la fracción de días cuando haya más de 13 ausentes.

1.35 Una compañía farmacéutica desea saber si un medicamento experimental tiene efecto sobre la presión sistólica de la sangre. A 15 pacientes seleccionados al azar se les aplicó el medicamento y, después de un tiempo suficiente para que el medicamento tuviera efecto, se registraron sus presiones sistólicas. Los datos aparecen a continuación:

172	140	123	130	115
148	108	129	137	161
123	152	133	128	142

a Calcule el valor de s usando la aproximación del rango.

b Calcule los valores de $\overline{y}$ y s para las 15 lecturas de presión sanguínea.

c Use el teorema de Tchebysheff (Ejercicio 1.32) para hallar valores de a y b tales que al menos 75% de las mediciones de presión sanguínea se encuentren entre a y b.

d ¿Funcionó el teorema de Tchebysheff? Es decir, use la información para hallar el porcentaje real de lecturas de presión sanguínea que están entre los valores de a y b hallados en el inciso c. ¿Este porcentaje real es mayor que 75%?

1.36 Una muestra aleatoria de 100 zorros fue examinada por un equipo de veterinarios para determinar el predominio de un parásito específico. Al contar el número de parásitos de este tipo específico, los veteri-

3. Los ejercicios precedidos de un asterisco son opcionales.

narios hallaron que 69 zorros no tenían los parásitos del tipo de interés, 17 no tenían un parásito del tipo en estudio y así sucesivamente. En la tabla siguiente se da un resumen de sus resultados.

Número de parásitos	0	1	2	3	4	5	6	7	8
Número de zorros	69	17	6	3	1	2	1	0	1

a Construya el histograma de frecuencia relativa para el número de parásitos por zorro.

b Calcule $\overline{y}$ y s para la información dada.

c ¿Qué fracción del conteo de parásitos cae dentro de 2 desviaciones estándar de la media? ¿Dentro de 3 desviaciones estándar? ¿Sus resultados concuerdan con el teorema de Tchebysheff (Ejercicio 1.32) y/o la regla empírica?

1.37 Estudios realizados indican que el agua potable suministrada por algunos viejos sistemas municipales de tubería con recubrimiento de plomo pueden contener niveles peligrosos de éste. Con base en datos presentados por Karalekas y colegas,[4] parece que la distribución de lecturas de contenido de plomo para especímenes individuales de agua tiene una media de .033 mg/L y desviación estándar de .10 mg/L. Explique por qué es obvio que las lecturas de contenido de plomo *no están* normalmente distribuidas.

1.38 En el Ejercicio 1.19 la media y la desviación estándar de la cantidad de cloroformo presente en fuentes de agua eran 34 y 53, respectivamente. Usted argumentó que las cantidades de cloroformo podrían no estar normalmente distribuidas. Use el teorema de Tchebysheff (Ejercicio 1.32) para describir la distribución de las cantidades de cloroformo en fuentes de agua.

4. P. C. Karalekas, Jr., C. R. Ryan y F. B. Taylor, "Control of Lead, Copper and Iron Pipe Corrosion in Boston", *American Water Works Journal* (febrero de 1983): 92.

CAPÍTULO 2

Probabilidad

2.1 Introducción

En nuestras conversaciones cotidianas, el término *probabilidad* es una medida de la creencia de que un evento futuro pueda ocurrir. Aceptamos esto como una interpretación significativa y práctica de probabilidad pero buscamos un concepto más claro de su contexto, de cómo se mide y cómo ayuda a hacer inferencias.

El concepto de probabilidad es necesario para trabajar con mecanismos físicos, biológicos o sociales que generan observaciones que no se pueden predecir con certeza. Por ejemplo, la presión sanguínea de una persona en un punto dado en el tiempo no se puede predecir con certeza y nunca sabemos la carga exacta que un puente resistirá antes de desplomarse en un río. Eventos de esa naturaleza no se pueden predecir con certeza, pero la frecuencia relativa con la que ocurren en una larga serie de intentos es a veces sorprendentemente estable. Los eventos que poseen esta propiedad reciben el nombre de eventos *aleatorios* o *estocásticos.*

Esta frecuencia relativa a largo plazo proporciona una medida intuitivamente significativa de nuestra creencia de que ocurrirá un evento aleatorio si se hace una observación futura. Es imposible, por ejemplo, predecir con certeza que una moneda normal caiga de cara en un solo tiro, pero estaríamos dispuestos a decir con un grado razonable de confianza que la fracción de caras en una larga serie de intentos sería muy cercana a .5. Que esta frecuencia relativa se use comúnmente como medida de la creencia en el resultado de un solo tiro es evidente cuando consideramos la probabilidad desde el punto de vista de un jugador. Éste arriesga dinero en un solo tiro de una moneda, no en una larga serie de tiros. La frecuencia relativa de que aparezca una cara en una larga serie de tiros, que un jugador llama probabilidad de una cara, le da una medida de la probabilidad de ganar en un solo tiro. Si la moneda no estuviera balanceada y diera 90% de caras en una larga serie de tiros, el jugador diría que la probabilidad de que salga una cara es .9 y tendría confianza en que saliera una cara en un solo tiro de la moneda.

El ejemplo anterior posee algunas analogías realistas y prácticas. En muchos aspectos todos somos jugadores. Una física investigadora apuesta tiempo y dinero en un proyecto de investigación y se interesa por su éxito en un solo tiro de su moneda simbólica. Del mismo modo, la inversión de capital en una nueva planta de manufacturas es un juego que representa un solo tiro de una moneda en el que el empresario tiene grandes esperanzas de éxito. La fracción de inversiones similares que sean exitosas, en una larga serie de intentos, es de interés para el empresario sólo hasta donde dé una medida de la creencia en el resultado exitoso de una sola inversión individual.

El concepto de frecuencia relativa de probabilidad, aunque intuitivamente significativo, no da una definición rigurosa de probabilidad. Se han propuesto muchos otros conceptos de probabilidad, incluyendo el de probabilidad subjetiva que permite que la probabilidad de un evento varíe dependiendo de la persona que efectúe la evaluación. No obstante, para nuestros fines, aceptamos una interpretación basada en la frecuencia relativa como medida significativa de nuestra idea de que ocurrirá un evento. Examinemos a continuación el enlace que la probabilidad hace entre observación e inferencia.

2.2 Probabilidad e inferencia

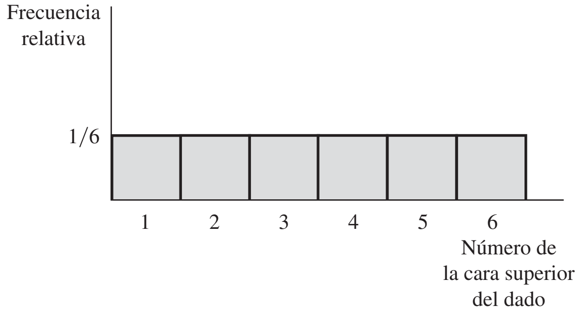
El papel que la probabilidad desempeña para hacer inferencias se estudiará con detalle después de sentar una base adecuada para la teoría de probabilidad. En este punto presentaremos un tratamiento elemental de esta teoría por medio de un ejemplo y una petición a su intuición.

El ejemplo seleccionado es similar al presentado en la Sección 1.4 pero más sencillo y menos práctico. Se escogió por la facilidad con la que podemos visualizar la población y muestra y porque da un mecanismo que produce observación, para el cual se construirá un modelo probabilístico en la Sección 2.3.

Considere a un jugador que desea hacer una inferencia respecto al balance de un dado. La población conceptual de interés es el conjunto de números que se generarían si el dado se tirara un número infinito de veces. Si el dado estuviera perfectamente balanceado, un sexto de las mediciones de esta población serían los números 1, un sexto de números 2, un sexto de números 3 y así sucesivamente. La correspondiente distribución de frecuencia se muestra en la Figura 2.1.

Si se usa el método científico, un jugador propone la hipótesis de *que el dado está balanceado* y busca observaciones de la naturaleza para contradecir la teoría, si es falsa. Se

FIGURA 2.1 Distribución de frecuencia para la población generada por un dado balanceado



selecciona una muestra de diez tiros de la población al tirar el dado diez veces; los diez tiros resultan en números 1. El jugador estima este resultado de la naturaleza con un ojo envidioso y concluye que su hipótesis no está de acuerdo con la naturaleza y por tanto que el dado *no* está balanceado.

El razonamiento empleado por el jugador identifica el papel que desempeña la probabilidad para hacer inferencias. El jugador rechazó su hipótesis (y concluyó que el dado no está balanceado) no porque sea *imposible* tirar diez números 1 en diez tiros de un dado balanceado sino porque esto es altamente *improbable.* Su evaluación de la probabilidad fue principalmente subjetiva. Esto es, el jugador puede no haber sabido cómo calcular la probabilidad de diez números 1 en diez tiros, pero tenía una impresión intuitiva de que este evento era muy poco probable si el dado estuviera balanceado. El punto a observar es que su decisión estuvo basada en la probabilidad de la muestra observada.

La necesidad de una teoría de probabilidad, que dará un método riguroso para hallar un número (o probabilidad) que estará de acuerdo con la frecuencia relativa real de que ocurra un evento en una larga serie de intentos, es evidente si imaginamos un resultado diferente para la muestra del jugador. Suponga, por ejemplo, que en lugar de diez números 1 él observó cinco números 1 junto con dos números 2, un 3, un 4 y un 6. ¿Este resultado es tan improbable que deberíamos rechazar nuestra hipótesis de que el dado está balanceado y concluir que el dado está cargado a favor de los números 1? Si debemos apoyarnos sólo en la experiencia e intuición para hacer nuestra evaluación, no es tan fácil decidir si la probabilidad de cinco números 1 en diez tiros es grande o pequeña. La probabilidad de tirar cuatro números 1 en diez tiros sería incluso más difícil de adivinar. No negaremos que resultados experimentales son en ocasiones obviamente inconsistentes con una hipótesis dada y llevan a su rechazo, pero muchos resultados experimentales caen en un área gris donde requerimos una evaluación rigurosa de la probabilidad de que ocurran. De hecho, no es difícil demostrar que evaluaciones intuitivas de probabilidades en ocasiones llevan a respuestas que están erróneas y resultan en inferencias incorrectas acerca de la población objetivo. Por ejemplo, si hay 20 personas en un cuarto, casi todos calcularían que es muy poco probable que hubiera dos o más personas con el mismo cumpleaños, pero, en ciertas suposiciones razonables, en el Ejemplo 2.18 demostraremos que la probabilidad de que eso ocurra es mayor que .4, número que es sorprendentemente grande para muchos.

Necesitamos una teoría de probabilidad que nos permita calcular la probabilidad (o una cantidad proporcional a la probabilidad) de observar resultados especificados, suponiendo que nuestro modelo hipotético sea correcto; este tema se desarrollará con detalle en los capítulos siguientes. Nuestra meta inmediata es presentar una introducción a la teoría de la probabilidad, que proporciona la base para la inferencia estadística moderna. Empezaremos por repasar alguna notación de conjuntos que usaremos para construir modelos probabilísticos para experimentos.

2.3 Un repaso de notación de conjuntos

Para continuar con un desarrollo ordenado de la teoría de probabilidad, necesitamos algunos conceptos básicos de teoría de conjuntos. Usaremos letras mayúsculas $A, B, C, \ldots$, para denotar conjuntos de puntos. Si los elementos del conjunto A son a_1, a_2 y a_3, escribiremos

$$A = \{a_1, a_2, a_3\}.$$

Denotemos con S el conjunto de todos los elementos en consideración; esto es, S es el *conjunto universal.* Para dos conjuntos cualesquiera A y B, diremos que A es un *subconjunto* de B, o A está contenido en B (denotado $A \subset B$), si todo punto en A también está en B. El *conjunto nulo*, o *vacío*, denotado por $\emptyset$, es el conjunto que no contiene puntos. Entonces, $\emptyset$ es un subconjunto de todo conjunto.

Los conjuntos y las relaciones entre conjuntos se pueden representar en forma conveniente con el uso de *diagramas de Venn.* El diagrama de Venn de la Figura 2.2 muestra dos conjuntos, A y B, del conjunto universal S. El conjunto A es el conjunto de todos los puntos dentro del triángulo; el conjunto B es el conjunto de todos los puntos dentro del círculo. Observe que en la Figura 2.2, $A \subset B$.

Considere ahora dos conjuntos arbitrarios de puntos. La *unión* de A y B, denotada por $A \cup B$, es el conjunto de todos los puntos en A o B o en ambos. Esto es, la unión de A y B contiene todos los puntos que están en al menos uno de los conjuntos. El diagrama de Venn de la Figura 2.3 muestra dos conjuntos A y B, donde A es el conjunto de puntos en el círculo

FIGURA **2.2**
Diagrama de Venn para $A \subset B$

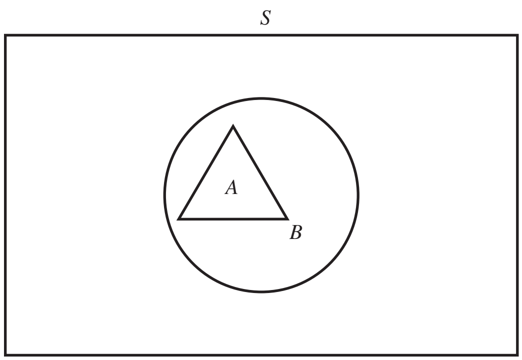


FIGURA **2.3**
Diagrama de Venn para $A \cup B$

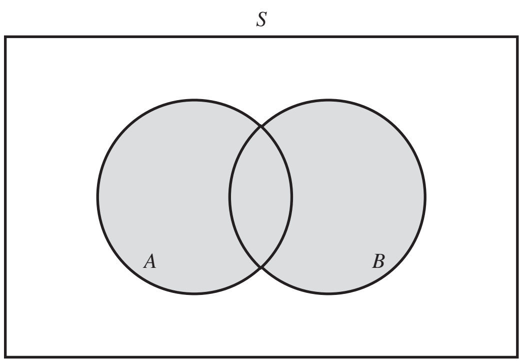
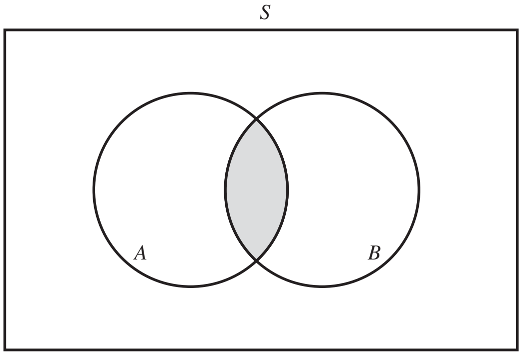


FIGURA **2.4**
Diagrama de Venn para AB



izquierdo y B es el conjunto de puntos en el círculo derecho. El conjunto $A \cup B$ es la región sombreada formada por todos los puntos dentro de cualquiera de los círculos (o de ambos). La palabra clave para expresar la unión de dos conjuntos es *o* (que significa A o B o ambos).

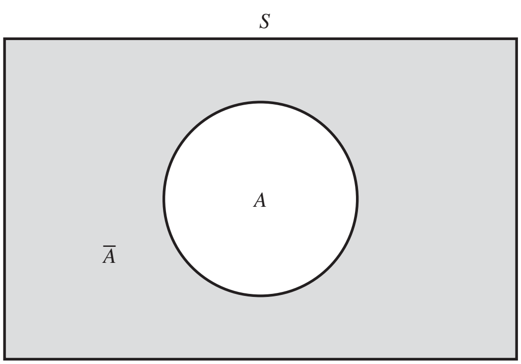
La *intersección* de A y B, denotada por $A \cap B$ o por AB, es el conjunto de todos los puntos en A y B. El diagrama de Venn de la Figura 2.4 muestra dos conjuntos A y B, con $A \cap B$ formado por los puntos en la región sombreada donde los dos conjuntos se traslapan. La palabra clave para expresar intersecciones es *y* (que significa A y B *simultáneamente*).

Si A es un subconjunto de S, entonces el *complemento* de A, denotado por $\overline{A}$, es el conjunto de puntos que están en S pero no en A. La Figura 2.5 es un diagrama de Venn que ilustra que el área sombreada en S pero no en A es $\overline{A}$. Observe que $A \cup \overline{A} = S$.

Se dice que dos conjuntos, A y B, son *disjuntos* o *mutuamente excluyentes*, si $A \cap B = \emptyset$. Esto es, los conjuntos mutuamente excluyentes no tienen puntos en común. El diagrama de Venn de la Figura 2.6 ilustra dos conjuntos A y B que son mutuamente excluyentes. Examinando la Figura 2.5 es fácil ver que, para cualquier conjunto A, A y $\overline{A}$ son mutuamente excluyentes.

Considere el problema de la Sección 2.2 de lanzar un dado y denote con S el conjunto de todas las posibles observaciones numéricas para un solo tiro de un dado. Esto es, $S = \{1, 2, 3, 4, 5, 6\}$. Sea $A = \{1, 2\}$, $B = \{1, 3\}$ y $C = \{2, 4, 6\}$. Entonces $A \cup B = \{1, 2, 3\}$, $A \cap B = \{1\}$ y $\overline{A} = \{3, 4, 5, 6\}$. Del mismo modo, observe que B y C son mutuamente excluyentes, mientras que A y C no lo son.

FIGURA **2.5**
Diagrama de Venn para $\overline{A}$



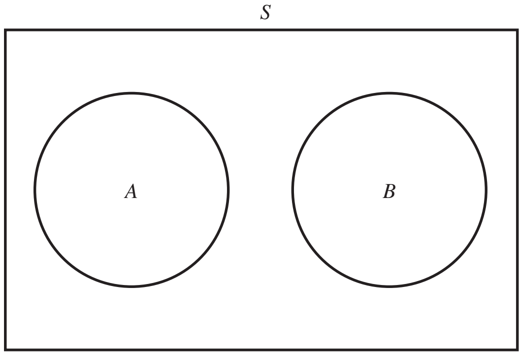


FIGURA 2.6
Diagrama de Venn para los conjuntos A y B mutuamente excluyentes

No trataremos de hacer aquí un repaso a fondo de álgebra de conjuntos, pero mencionamos cuatro igualdades de considerable importancia. Éstas son las *leyes distributivas*, dadas por

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

y las *leyes* de *DeMorgan:*

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B} \quad \text{y} \quad \overline{(A \cup B)} = \overline{A} \cap \overline{B}.$$

En la siguiente sección continuaremos con una exposición elemental de la teoría de probabilidad.

Ejercicios

2.1 Suponga que una familia contiene dos hijos de edades diferentes y estamos interesados en el género de estos niños. Denotemos con F que una hija es mujer y M que el hijo es hombre y denote con un par, por ejemplo FM, que el hijo de mayor edad es la niña y el más joven es el niño. Hay cuatro puntos en el conjunto S de posibles observaciones:

$$S = \{FF,\ FM,\ MF,\ MM\}.$$

Denote con A el subconjunto de posibilidades que no contenga hombres; B, el subconjunto que contiene dos hombres; y C, el subconjunto que contenga al menos un hombre. Indique los elementos de A, B, C, $A \cap B$, $A \cup B$, $A \cap C$, $A \cup C$, $B \cap C$, $B \cup C$ y $C \cap \overline{B}$.

2.2 Suponga que A y B son dos eventos. Escriba las expresiones que contengan uniones, intersecciones y complementos que describan lo siguiente:

a Ocurren ambos eventos.

b Al menos uno ocurre.

c Ninguno ocurre.

d Exactamente uno ocurre.

2.3 Trace diagramas de Venn para verificar las leyes de DeMorgan. Esto es, para dos conjuntos cualesquiera A y B, $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$ y $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$.

2.4 Si A y B son dos conjuntos, dibuje diagramas de Venn para verificar lo siguiente:

a $A = (A \cap B) \cup (A \cap \overline{B})$.
b Si $B \subset A$ entonces $A = B \cup (A \cap \overline{B})$.

2.5 Consulte el Ejercicio 2.4. Use las identidades $A = A \cap S$ y $S = B \cup \overline{B}$ y una ley distributiva para demostrar que

a $A = (A \cap B) \cup (A \cap \overline{B})$.
b Si $B \subset A$ entonces $A = B \cup (A \cap \overline{B})$.
c Además, demuestre que $(A \cap B)$ y $(A \cap \overline{B})$ son mutuamente excluyentes y que, por tanto, A es la unión de dos conjuntos mutuamente excluyentes, $(A \cap B)$ y $(A \cap \overline{B})$
d También demuestre que B y $(A \cap \overline{B})$ son mutuamente excluyentes y, si $B \subset A$, A es la unión de dos conjuntos mutuamente excluyentes, B y $(A \cap \overline{B})$.

2.6 Suponga que se tiran dos dados y que se observan los números de las caras superiores. Denotemos con S el conjunto de todos los pares posibles que se pueden observar. [Estos pares se pueden indicar, por ejemplo, si con (2, 3) se denota que un 2 se ha observado en el primer dado y un 3 en el segundo.]

a Defina los siguientes subconjuntos de S:

A: el número en el segundo dado es par.

B: la suma de los dos números es par.

C: al menos un número del par es impar.

b Indique los puntos en A, $\overline{C}$, $A \cap B$, $A \cap \overline{B}$, $\overline{A} \cup B$ y $\overline{A} \cap C$.

2.7 Un grupo de cinco solicitantes para un par de trabajos idénticos está formado por tres hombres y dos mujeres. El empleador ha de seleccionar dos de los cinco solicitantes para los trabajos. Denote con S el conjunto de todos los resultados posibles para la selección del empleador. Denote con A al subconjunto de resultados correspondientes a la selección de dos hombres y con B al subconjunto correspondiente a la selección de al menos una mujer. Indique los resultados en A, $\overline{B}$, $A \cup B$, $A \cap B$ y $A \cap \overline{B}$. Denote los hombres y mujeres diferentes con M_1, M_2, M_3 y W_1, W_2, respectivamente.)

2.8 De una encuesta de 60 estudiantes que asisten a clase en una universidad, se encontró que 9 vivían fuera del campus, 36 eran pasantes y 3 eran pasantes que vivían fuera del campus. Encuentre el número de estos estudiantes que

a eran pasantes, vivían fuera del campus o ambos.
b eran pasantes que vivían en el campus.
c eran graduados que vivían en el campus.

2.4 Un modelo probabilístico para un experimento: el caso discreto

En la Sección 2.2 nos referimos al *experimento* de lanzar un dado cuando observamos el número que aparecía en la cara superior. Usaremos el término *experimento* para incluir observaciones obtenidas de situaciones incontrolables por completo (por ejemplo observaciones en el precio diario de una acción en particular) así como aquellas hechas en condiciones controladas de laboratorio. Tenemos la siguiente definición:

DEFINICIÓN **2.1**

Un *experimento* es el proceso por medio del cual se hace una observación.

Ejemplos de experimentos incluyen el tiro de monedas y dados, medir el cociente de inteligencia (IQ) de una persona o determinar el número de bacterias por centímetro cúbico en una porción de alimento procesado.

Cuando se realiza un experimento, puede terminar en uno o más resultados que se llaman *eventos*. En nuestras exposiciones, los eventos se denotarán con letras mayúsculas. Si el experimento consiste en contar el número de bacterias en una porción de alimento, algunos eventos de interés podrían ser

A: Exactamente 110 bacterias están presentes.
B: Más de 200 bacterias están presentes.
C: El número de bacterias presentes está entre 100 y 300.

Algunos eventos asociados con un solo tiro de un dado balanceado son los siguientes:

A: Observar un número impar.
B: Observar un número menor que 5.
C: Observar un 2 o un 3.
E_1: Observar un 1.
E_2: Observar un 2.
E_3: Observar un 3.
E_4: Observar un 4.
E_5: Observar un 5.
E_6: Observar un 6.

Usted puede ver que hay una clara diferencia entre algunos de los eventos asociados con el experimento de tirar un dado. Por ejemplo, si observó el evento A (un número impar), al mismo tiempo habrá observado E_1, E_3 o E_5. Así, el evento A, que se puede descomponer en otros tres eventos, se denomina *evento compuesto*. En contraste, los eventos E_1, E_2, E_3, E_4, E_5 y E_6 no se pueden descomponer y reciben el nombre de *eventos simples*. Un evento simple puede ocurrir sólo en un modo, mientras que un evento compuesto puede ocurrir en más de una forma distinta.

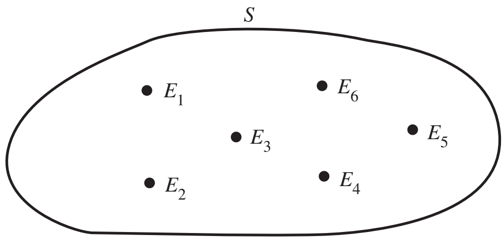
Ciertos conceptos de teoría de conjuntos son útiles para expresar las relaciones entre diversos eventos asociados con un experimento. Debido a que los conjuntos son grupos de puntos, asociamos un punto distinto, llamado *punto muestral*, con todos y cada uno de los eventos simples asociados con un experimento.

DEFINICIÓN **2.2**

Un *evento simple* no se puede descomponer. Cada evento simple corresponde a un y sólo un *punto muestral*. La letra E con un subíndice se empleará para denotar un evento simple o el correspondiente punto muestral.

Así, podemos considerar un evento simple como un conjunto formado por un solo punto, es decir, el solo punto muestral asociado con el evento.

FIGURA **2.7** Diagrama de Venn para el espacio muestral asociado con el experimento de tirar un dado



DEFINICIÓN **2.3**

El *espacio muestral* asociado con un experimento es el conjunto formado por todos los posibles puntos muestrales. Un espacio muestral estará denotado por S.

Fácilmente podemos ver que el espacio muestral S asociado con el experimento de lanzar un dado está formado por seis puntos muestrales correspondientes a los seis eventos simples E_1, E_2, E_3, E_4, E_5 y E_6. Esto es, $S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$. Un diagrama de Venn que exhibe el espacio muestral para el experimento de lanzar un dado se da en la Figura 2.7.

Para el ejemplo de microbiología de contar bacterias en un espécimen de alimento, hagamos que E_0 corresponda a observar 0 bacterias, E_1 a observar 1 bacteria y así sucesivamente. Entonces el espacio muestral es

$$S = \{E_0, E_1, E_2, \dots\}$$

porque no se puede excluir ningún número entero de bacterias como posible resultado.

Ambos espacios muestrales que examinamos tienen la propiedad de que están formados ya sea por un número finito o uno contable de puntos muestrales. En el ejemplo de lanzar un dado hay seis (un número finito) puntos muestrales. El número de puntos muestrales asociado con el experimento de contar bacterias es infinito, pero el número de puntos muestrales distintos se puede poner en correspondencia biunívoca con los enteros (es decir, el número de puntos muestrales es contable). Se dice que estos espacios muestrales son discretos.

DEFINICIÓN **2.4**

Un *espacio muestral discreto* es aquel que está formado ya sea por un número finito o uno contable de puntos muestrales distintos.

Cuando un experimento se realiza en una sola vez, se observará un y sólo un evento simple. Por ejemplo, si se lanza un dado y se observa un 1, no se puede al mismo tiempo observar un 2. Así, el punto muestral simple E_1 asociado con observar un 1 y el punto muestral simple E_2 asociado con observar un 2 son distintos, y los conjuntos $\{E_1\}$ y $\{E_2\}$ son conjuntos mutuamente excluyentes. Por tanto, los eventos E_1 y E_2 son mutuamente excluyentes. Del mismo modo, todos los eventos simples distintos corresponden a conjuntos mutuamente excluyentes de eventos simples y son por tanto eventos mutuamente excluyentes.

Para experimentos con espacios muestrales discretos, los *eventos compuestos* se pueden ver como grupos (conjuntos) de puntos muestrales o bien, de manera equivalente, como uniones de los conjuntos de puntos muestrales simples correspondientes a los eventos simples apropiados. Por ejemplo, el evento A de lanzar un dado (observar un número impar) ocurrirá

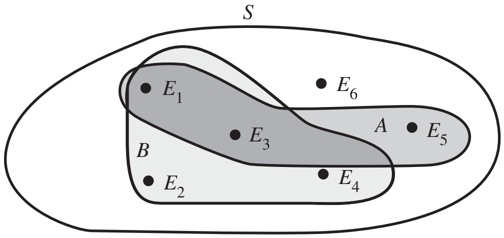


FIGURA **2.8** Diagrama de Venn para el experimento de lanzar un dado

si y sólo si ocurre uno de los eventos simples E_1, E_3 o E_5. Así,

$$A = \{E_1,\ E_3,\ E_5\} \quad \text{o} \quad A = E_1 \cup E_3 \cup E_5.$$

Del mismo modo, B (observar un número menor que 5) se puede escribir como

$$B = \{E_1,\ E_2,\ E_3,\ E_4\} \quad \text{o} \quad B = E_1 \cup E_2 \cup E_3 \cup E_4.$$

La regla para determinar cuáles eventos simples incluir en un evento compuesto es muy precisa. *Un evento simple E_i se incluye en el evento A si y sólo si A ocurre siempre que ocurra E_i.*

DEFINICIÓN **2.5**

Un *evento* en un espacio muestral discreto S es un conjunto de puntos muestrales, es decir, cualquier subconjunto de S.

La Figura 2.8 da un diagrama de Venn que representa el espacio muestral y los eventos A (observar un número impar) y B (observar un número menor que 5) para el experimento de lanzar un dado. Observe que es fácil visualizar la relación entre eventos si se usa un diagrama de Venn.

Por la Definición 2.5, cualquier evento en un espacio muestral S es un subconjunto de S. En el ejemplo relacionado con determinar el número de bacterias en una porción de alimento, el evento B (el número de bacterias es mayor que 200) se puede expresar como

$$B = \{E_{201},\ E_{202},\ E_{203}, \ldots\},$$

donde E_i denota el evento simple de que hay i bacterias presentes en la muestra de alimento e $i = 0, 1, 2, \ldots$

Un modelo probabilístico para un experimento con un espacio muestral discreto se puede construir al asignar una probabilidad numérica a cada evento simple del espacio muestral S. Seleccionaremos este número, una medida de nuestra creencia en que ocurrirá el evento en una sola repetición del experimento, de forma tal que será consistente con el concepto de frecuencia relativa de probabilidad. Aun cuando la frecuencia relativa no da una definición rigurosa de probabilidad, cualquier definición aplicable al mundo real debe concordar con nuestra noción intuitiva de las frecuencias relativas de eventos.

Al analizar el concepto de frecuencia relativa de probabilidad, vemos que deben cumplirse tres condiciones.

1. La frecuencia relativa de que ocurra cualquier evento debe ser mayor o igual a cero. Una frecuencia relativa negativa no tiene sentido.

2. La frecuencia relativa de todo el espacio muestral S debe ser la unidad. Debido a que todo posible resultado del experimento es un punto en S, se deduce que S debe ocurrir cada vez que se realice el experimento.
3. Si dos eventos son mutuamente excluyentes, la frecuencia relativa de la unión de ambos es la suma de sus respectivas frecuencias relativas. (Por ejemplo, si el experimento de lanzar un dado balanceado da un 1 en 1/6 de los tiros, debe dar un 1 o un 2 con $1/6 + 1/6 = 1/3$ de tiros.)

Estas tres condiciones forman la base de la siguiente definición de probabilidad.

DEFINICIÓN 2.6

Suponga que S es un espacio muestral asociado con un experimento. A todo evento A en S (A es el subconjunto de S) le asignamos un número, $P(A)$, llamado *probabilidad* de A, de modo que se cumplen los siguientes axiomas:

Axioma 1: $P(A) \geq 0$.
Axioma 2: $P(S) = 1$.
Axioma 3: Si $A_1, A_2, A_3, \ldots$ forman una secuencia de eventos por pares mutuamente excluyentes en S (es decir, $A_i \cap A_j = \emptyset$ si $i \neq j$), entonces

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

Podemos fácilmente demostrar que el Axioma 3, que está expresado en términos de una sucesión infinita de eventos, implica una propiedad similar para una sucesión finita. Específicamente, si $A_1, A_2, \ldots, A_n$ son eventos mutuamente excluyentes por pares, entonces

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

Observe que la definición expresa sólo las condiciones que una asignación de probabilidades debe satisfacer: no nos dice cómo asignar probabilidades específicas a los eventos. Por ejemplo, suponga que una moneda ha dado 800 "caras" en 1000 tiros previos. Considere el experimento de un tiro más de la misma moneda. Hay dos posibles resultados, cara o cruz, y por tanto dos eventos simples. La definición de probabilidad nos permite asignar a estos eventos simples dos números no negativos cualesquiera que sumen 1. Por ejemplo, cada evento simple podría tener la probabilidad 1/2. En vista de la historia de esta moneda, no obstante, podría ser más razonable asignar una probabilidad más cercana a .8 al resultado donde aparece una cara. Las asignaciones específicas de probabilidades deben ser consistentes con la realidad si el modelo probabilístico ha de servir a un propósito útil.

Para espacios muestrales discretos, es suficiente asignar probabilidades a cada evento simple. Si se usa un dado balanceado para el ejemplo de lanzar un dado, parece razonable suponer que todos los eventos simples tendrían la misma frecuencia relativa a la larga. Asignaremos una probabilidad de 1/6 a cada evento simple: $P(E_i) = 1/6$, para $i = 1, 2, \ldots, 6$. Esta asignación de probabilidades está de acuerdo con el Axioma 1. Para ver que el Axioma 2 se satisfaga, escribamos

$$P(S) = P(E_1 \cup E_2 \cup \ldots \cup E_6) = P(E_1) + P(E_2) + \cdots + P(E_6) = 1.$$

La segunda igualdad es consecuencia de que el Axioma 3 debe cumplirse. El Axioma 3 nos dice que podemos calcular la probabilidad de cualquier evento al sumar las probabilidades de los eventos simples que contiene (recuerde que los eventos simples distintos son mutuamente excluyentes). El evento A se definió como "observar un número impar". Por tanto,

$$P(A) = P(E_1 \cup E_3 \cup E_5) = P(E_1) + P(E_3) + P(E_5) = 1/2.$$

EJEMPLO **2.1** Un fabricante tiene cinco terminales de computadora aparentemente idénticas listas para ser enviadas. Sin que él lo sepa, dos de las cinco están defectuosas. Un pedido en particular solicita dos de las terminales y el pedido se surte seleccionando al azar dos de las cinco de que se dispone.

a Indique el espacio muestral para el experimento.

b Denote con A el evento de que el pedido se surta con dos terminales no defectuosas. Indique los puntos muestrales en A.

c Construya un diagrama de Venn para el experimento que ilustre el evento A.

d Asigne probabilidades a los eventos simples en tal forma que se use la información acerca del experimento y se satisfagan los axiomas de la Definición 2.6.

e Encuentre la probabilidad del evento A.

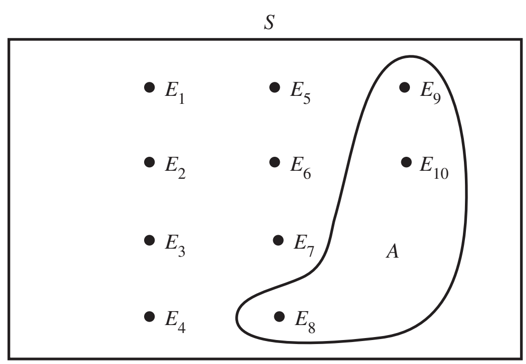
Solución **a** Marque con D_1 y D_2 las dos terminales defectuosas y con G_1, G_2 y G_3 las tres terminales buenas. Cualquier punto muestral estará formado por una lista de las dos terminales seleccionadas para envío. Los eventos simples pueden denotarse con

$$\begin{array}{llll} E_1 = \{D_1, D_2\}, & E_5 = \{D_2, G_1\}, & E_8 = \{G_1, G_2\}, & E_{10} = \{G_2, G_3\}. \\ E_2 = \{D_1, G_1\}, & E_6 = \{D_2, G_2\}, & E_9 = \{G_1, G_3\}, & \\ E_3 = \{D_1, G_2\}, & E_7 = \{D_2, G_3\}, & & \\ E_4 = \{D_1, G_3\}, & & & \end{array}$$

Así, hay diez puntos muestrales en S, y $S = \{E_1, E_2, \ldots, E_{10}\}$.

b Evento $A = \{E_8, E_9, E_{10}\}$.

c



d Debido a que las terminales se seleccionan al azar, cualquier par de terminales es tan probable de ser seleccionado como cualquier otro. Entonces, $P(E_i) = 1/10$, para $i = 1, 2, \ldots, 10$, es una asignación razonable de probabilidades.

e Como $A = E_8 \cup E_9 \cup E_{10}$, el Axioma 3 implica que

$$P(A) = P(E_8) + P(E_9) + P(E_{10}) = 3/10. \qquad \blacksquare$$

La siguiente sección contiene una descripción axiomática del método para calcular $P(A)$ que acabamos de usar.

Antes de continuar, tomemos nota de que hay experimentos para los cuales el espacio muestral no es contable y por lo tanto no es discreto. Suponga, por ejemplo, que el experimento consiste en medir el nivel de glucosa en sangre de un paciente diabético. El espacio muestral para este experimento contendría un intervalo de números reales y cualquiera de estos intervalos contiene un número incontable de valores. Por tanto, el espacio muestral no es discreto. En el Capítulo 4 examinaremos situaciones como ésta; el resto de este capítulo está dedicado a desarrollar métodos para calcular las probabilidades de eventos definidos en espacios muestrales discretos.

Ejercicios

2.9 El tipo de sangre de todas las personas es A, B, AB u O. Además, cada persona tiene también el factor Rhesus (Rh) (+) o (−). Un técnico médico registra el tipo de sangre de una persona y el factor Rh. Indique el espacio muestral para este experimento.

2.10 Las proporciones de fenotipos sanguíneos A, B, AB y O, en la población de todos los caucásicos en Estados Unidos son aproximadamente .41, .10, .04 y .45, respectivamente. Un solo caucásico se selecciona al azar de la población.

a Indique el espacio muestral para este experimento.

b Haga uso de la información dada antes para asignar probabilidades a cada uno de los eventos simples.

c ¿Cuál es la probabilidad de que la persona seleccionada al azar tenga tipo de sangre A o tipo AB?

2.11 Un espacio muestral está formado por cinco eventos simples, E_1, E_2, E_3, E_4 y E_5.

a Si $P(E_1) = P(E_2) = 0.15$, $P(E_3) = 0.4$ y $P(E_4) = 2P(E_5)$, halle las probabilidades de E_4 y E_5.

b Si $P(E_1) = 3P(E_2) = 0.3$, encuentre las probabilidades de los eventos simples restantes si se sabe que son igualmente probables.

2.12 Un vehículo que llega a un crucero puede dar vuelta a la derecha, a la izquierda o continuar de frente. El experimento consiste en observar el movimiento de un solo vehículo por el crucero.

a Indique el espacio muestral para este experimento.

b Suponiendo que todos los puntos muestrales son igualmente probables, encuentre la probabilidad de que el vehículo dé vuelta.

2.13 Los estadounidenses pueden ser bastante suspicaces, en especial cuando se trata de conspiraciones del gobierno. Sobre la pregunta de si la Fuerza Aérea de Estados Unidos tiene oculta la prueba de la existencia de vida inteligente en otros planetas, las proporciones de estadounidenses con opiniones que varían se dan en la tabla.

Opinión	Proporción
Muy probable	.24
Poco probable	.24
Improbable	.40
Otra	.12

Suponga que se selecciona un estadounidense y que se registra su opinión.

a ¿Cuáles son los eventos simples para este experimento?

b Todos los eventos simples que usted dio en el inciso a, ¿son igualmente probables? Si no es así, ¿cuáles son las probabilidades que deben asignarse a cada uno?

c ¿Cuál es la probabilidad de que la persona seleccionada encuentre al menos un poco probable que la Fuerza Aérea esté ocultando información acerca de vida inteligente en otros planetas?

2.14 Una encuesta clasificó a gran número de adultos de acuerdo con si se les diagnosticó la necesidad de usar lentes para corregir su visión de lectura o si ya usan lentes cuando leen. Las proporciones que caen en las cuatro categorías resultantes se dan en la tabla siguiente:

	Usa lentes para leer	
Necesita lentes	Sí	No
Sí	.44	.14
No	.02	.40

Si se selecciona un solo adulto del grupo grande, encuentre las probabilidades de los eventos definidas a continuación. El adulto

a necesita lentes,

b necesita lentes pero no los usa,

c usa lentes los necesite o no.

2.15 Una empresa de exploración petrolera encuentra petróleo o gas en 10% de sus perforaciones. Si la empresa perfora dos pozos, los cuatro posibles eventos simples y tres de sus probabilidades asociadas se dan en la tabla siguiente. Encuentre la probabilidad de que la compañía encuentre petróleo o gas

a en la primera perforación pero no en la segunda,

b en al menos una de las dos perforaciones.

Evento simple	Resultado de la primera perforación	Resultado de la segunda perforación	Probabilidad
E_1	Encuentra (petróleo o gas)	Encuentra (petróleo o gas)	.01
E_2	Encuentra	No encuentra	?
E_3	No encuentra	Encuentra	.09
E_4	No encuentra	No encuentra	.81

2.16 De los voluntarios que entran en un centro de sangre, 1 en 3 tienen sangre O^+, 1 en 15 tienen O^-, 1 en 3 tienen A^+ y 1 en 16 tienen A^-. El nombre de una persona que previamente ha donado sangre se selecciona de los registros del centro. ¿Cuál es la probabilidad de que la persona seleccionada tenga

a tipo de sangre O^+?
b tipo de sangre O?
c tipo de sangre A?
d ni tipo A ni tipo O de sangre?

2.17 Los trenes de aterrizaje hidráulicos que salen de una planta de reparación de aviones se inspeccionan para ver si tienen defectos. Registros históricos indican que 8% tienen defectos sólo en ejes, 6% tienen defectos sólo en bujes y 2% tienen defectos en ejes y bujes. Uno de los trenes hidráulicos se selecciona al azar. ¿Cuál es la probabilidad de que el conjunto tenga

a un buje defectuoso?
b un eje o buje defectuoso?
c exactamente uno de los dos tipos de defecto?
d ningún tipo de defecto?

2.18 Suponga que dos monedas balanceadas se tiran al aire y que se observan las caras superiores.

a Indique los puntos muestrales para este experimento.
b Asigne una probabilidad razonable a cada punto muestral. (¿Los puntos muestrales son igualmente probables?)
c Denote con A el evento de que *exactamente* se vea una cara y con B el evento de que se vea *al menos* una cara. Indique los puntos muestrales en A y B.
d De su respuesta al inciso c, encuentre $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$ y $P(\overline{A} \cup B)$.

2.19 Una oficina de finanzas solicita suministros de papel de uno de tres vendedores, V_1, V_2 o V_3. Los pedidos han de colocarse en dos días sucesivos, un pedido por día. Así, (V_2, V_3) podría denotar que el vendedor V_2 obtiene el pedido en el primer día y el vendedor V_3 obtiene el pedido en el segundo día.

a Indique los puntos muestrales en este experimento de solicitar papel en dos días sucesivos.
b Suponga que los vendedores se seleccionan al azar cada día y se asigna probabilidad a cada punto muestral.
c Denote con A el evento de que el mismo vendedor obtenga ambos pedidos y B el evento de que V_2 obtenga al menos un pedido. Encuentre $P(A)$, $P(B)$, $P(A \cup B)$ y $P(A \cap B)$ al suponer las probabilidades de los puntos muestrales en estos eventos.

***2.20** El siguiente juego se jugó en un conocido programa de televisión. El presentador mostró tres cortinas grandes a una concursante. Detrás de una de las cortinas estaba un bonito premio (quizá un auto nuevo) y detrás de las otras dos cortinas había premios sin valor (falsos). A la concursante se le pidió escogiera una cortina. Si las cortinas eran identificadas por sus premios, podrían estar marcadas G, D_1 y D_2 (buen premio, falso1 y falso2). Entonces, el espacio muestral para la selección de los concursantes es $S = \{G, D_1, D_2\}$.[1]

a Si la concursante no tiene idea de cuáles cortinas ocultan los premios y selecciona una cortina al azar, asigne probabilidades razonables a los eventos sencillos y calcule la probabilidad de que la concursante seleccione la cortina que oculta el premio bueno.
b Antes de mostrar a la concursante lo que estaba detrás de la cortina inicialmente escogida, el presentador del juego abriría una de las cortinas y mostraría a la concursante uno de los premios sin valor (siempre podría hacer esto porque sabe cuál cortina oculta el premio bueno). Luego ofreció a la concursante la opción de cambiar de la cortina inicialmente seleccionada a la otra cortina restante no abierta. ¿Cuál estrategia lleva al máximo la probabilidad que tiene la concursante de ganar el premio bueno: quedarse con la opción inicial o cambiar a la otra cortina? Al contestar la siguiente secuencia

1. Los ejercicios precedidos por un asterisco son opcionales.

de preguntas, usted descubrirá que, quizá de una manera sorprendente, esta pregunta puede ser contestada si se considera sólo el espacio muestral citado antes y usar las probabilidades que asignó para contestar el inciso a.

i Si la concursante escoge quedarse con su selección inicial, gana el premio bueno si y sólo si inicialmente escogió la cortina G. Si se queda con su selección inicial, ¿cuál es la probabilidad de que gane el premio bueno?

ii Si el presentador le muestra uno de los premios sin valor y ella cambia a la otra cortina sin abrir, ¿cuál será el resultado si inicialmente ella hubiera seleccionado G?

iii Conteste la pregunta del inciso ii si ella inicialmente seleccionó uno de los premios sin valor.

iv Si la concursante cambia su selección inicial (como resultado de haberle sido mostrado uno de los premios sin valor), ¿cuál es la probabilidad de que la concursante gane el premio bueno?

v ¿Cuál estrategia lleva al máximo la probabilidad de la concursante de ganar el premio bueno: quedarse con su selección inicial o cambiar a la otra cortina?

***2.21** Si A y B son eventos, use el resultado obtenido en el Ejercicio 2.5(a) y los axiomas de la Definición 2.6 para demostrar que

$$P(A) = P(A \cap B) + P(A \cap \overline{B}).$$

***2.22** Si A y B son eventos y $B \subset A$, use el resultado obtenido en el Ejercicio 2.5(b) y los axiomas de la Definición 2.6 para demostrar que

$$P(A) = P(B) + P(A \cap \overline{B}).$$

2.23 Si A y B son eventos y $B \subset A$, ¿por qué es “obvio” que $P(B) \leq P(A)$?

2.24 Use el resultado del Ejercicio 2.22 y los axiomas de la Definición 2.6 para demostrar el resultado “obvio” del Ejercicio 2.23.

2.5 Cálculo de la probabilidad de un evento: el método de punto muestral

La probabilidad de un evento definido en un espacio muestral que contenga un conjunto de puntos muestrales finito o que se pueda contar aun cuando sea infinito, se puede calcular en dos formas: el método del *punto muestral* y el de la *composición de evento*. Ambos métodos usan el modelo de espacio muestral, pero difieren en la *secuencia* de pasos necesaria para obtener una solución y en las herramientas que se usan. La separación de los dos procedimientos puede no ser agradable para el teórico que busca la unidad, pero puede ser sumamente útil para un principiante que trate de hallar la probabilidad de un evento. En esta sección consideramos el método de punto muestral. El método de composición de eventos requiere resultados adicionales y se presentará en la Sección 2.9.

El método de punto muestral se compendia en la Sección 2.4. Los pasos siguientes se usan para hallar la probabilidad de un evento:

1. Defina el experimento y determine con toda claridad cómo describir un evento simple.
2. Indique los eventos simples asociados con el experimento y pruebe cada uno para asegurarse que no se pueden descomponer. Esto define el espacio muestral S.
3. Asigne probabilidades razonables a los puntos muestrales en S, asegurándose de que $P(E_i) \geq 0$ y $\sum P(E_i) = 1$.
4. Defina el evento de interés, A, como un conjunto específico de puntos muestrales. (Un punto muestral está en A si A ocurre cuando se presenta el punto muestral. Pruebe *todos* los puntos muestrales en S para identificar los que estén en A.)
5. Encuentre $P(A)$ al sumar las probabilidades de los puntos muestrales en A.

Ilustraremos estos pasos con tres ejemplos.

EJEMPLO **2.2** Considere el problema de seleccionar dos solicitantes para un trabajo, de un grupo de cinco solicitantes e imagine que éstos varían en grado de competencia, 1 siendo el mejor, 2 el segundo mejor y así sucesivamente para 3, 4 y 5. Estas clasificaciones son por supuesto desconocidas para el empleador. Defina dos eventos A y B como:

A: El empleador selecciona al mejor y a uno de los dos solicitantes menos aptos (1 y 4 o 1 y 5).

B: El empleador selecciona al menos uno de los dos mejores.

Encuentre las probabilidades de estos eventos.

Solución Los pasos son como sigue:

1. El experimento comprende seleccionar al azar dos solicitantes de entre cinco. Denote la selección de los solicitantes 3 y 5 por $\{3, 5\}$.
2. Los diez eventos simples, con $\{i, j\}$ denotando la selección de los solicitantes i y j, son
3. Una selección aleatoria de dos de entre cinco da a cada par una probabilidad igual de

$$\begin{array}{llll} E_1: \{1, 2\}, & E_5: \{2, 3\}, & E_8: \{3, 4\}, & E_{10}: \{4, 5\}. \\ E_2: \{1, 3\}, & E_6: \{2, 4\}, & E_9: \{3, 5\}, & \\ E_3: \{1, 4\}, & E_7: \{2, 5\}, & & \\ E_4: \{1, 5\}, & & & \end{array}$$

selección. Por tanto, asignaremos a cada punto muestral la probabilidad 1/10. Esto es,

$$P(E_i) = 1/10 = .1, \qquad i = 1, 2, \ldots, 10.$$

4. Al comprobar los puntos muestrales vemos que B se presenta siempre que se presentan E_1, E_2, E_3, E_4, E_5, E_6 o E_7. En consecuencia, estos puntos muestrales están incluidos en B.

5. Por último, $P(B)$ es igual a la suma de las probabilidades de los puntos muestrales en B, o sea

$$P(B) = \sum_{i=1}^{7} P(E_i) = \sum_{i=1}^{7} .1 = .7.$$

Del mismo modo, vemos que el evento $A = E_3 \cup E_4$ y que $P(A) = .1 + .1 = .2$. ■

La solución de éste y otros problemas semejantes sería de importancia para el director de personal de una empresa.

EJEMPLO **2.3** Una moneda balanceada se lanza tres veces al aire. Calcule la probabilidad de que exactamente dos de los tres tiros resulten en caras.

Solución Los cinco pasos del método de punto muestral son como sigue:

1. El experimento consiste en observar los resultados (caras o cruces) para cada uno de los tres tiros de una moneda. Un evento simple para este experimento se puede simbolizar con una secuencia de tres letras H y T que representan caras y colas, respectivamente. La primera letra de la secuencia representa la observación de la primera moneda. La segunda letra representa la observación de la segunda moneda, y así sucesivamente.
2. Los ocho eventos simples en S son

$$\begin{array}{llll} E_1: HHH, & E_3: HTH, & E_5: HTT, & E_7: TTH, \\ E_2: HHT, & E_4: THH, & E_6: THT, & E_8: TTT. \end{array}$$

3. Como la moneda está balanceada, esperaríamos que los eventos simples fueran exactamente probables; esto es,

$$P(E_i) = 1/8, \qquad i = 1, 2, \ldots, 8.$$

4. El evento de interés, A, es el evento que resulta exactamente en que dos de los tiros sean caras. Un examen de los puntos muestrales verifica que

$$A = \{E_2, E_3, E_4\}.$$

5. Por último,

$$P(A) = P(E_2) + P(E_3) + P(E_4) = 1/8 + 1/8 + 1/8 = 3/8.$$ ■

Aun cuando los puntos muestrales de los espacios muestrales asociados con los Ejemplos 2.2 y 2.3 son igualmente probables, es importante darse cuenta que los puntos muestrales no necesitan ser igualmente probables. Veamos ahora un ejemplo para ilustrar este punto.

EJEMPLO **2.4** Las probabilidades son dos a uno de que, cuando A y B juegan tenis, A gane. Suponga que A y B juegan dos partidos. ¿Cuál es la probabilidad de que A gane al menos un partido?

Solución

1. El experimento consiste en observar al ganador (A o B) de cada uno de los dos partidos. Denote con AB el evento de que el jugador A gane el primer duelo y el jugador B gane el segundo.
2. El espacio muestral para el experimento consta de cuatro puntos muestrales:

$$E_1: AA, \quad E_2: AB, \quad E_3: BA, \quad E_4: BB$$

3. Como A tiene una mejor probabilidad de gana cualquier partido, no parece apropiado asignar iguales probabilidades a estos puntos muestrales. Como se verá en la Sección 2.9, en ciertas condiciones es razonable hacer la siguiente asignación de probabilidades:

$$P(E_1) = 4/9, \qquad P(E_2) = 2/9, \qquad P(E_3) = 2/9, \qquad P(E_4) = 1/9.$$

Observe que, aun cuando las probabilidades asignadas a los eventos simples no son todas iguales, $P(E_i) \geq 0$, para $i = 1, 2, 3, 4$ y $\sum_S P(E_i) = 1$.

4. El evento de interés es que A gana al menos un juego. Así, si denotamos el evento de interés como C, fácilmente se ve que

$$C = E_1 \cup E_2 \cup E_3.$$

5. Por último,

$$P(C) = P(E_1) + P(E_2) + P(E_3) = 4/9 + 2/9 + 2/9 = 8/9. \qquad \blacksquare$$

El método de punto muestral para resolver un problema de probabilidades es directo y poderoso y en algunos aspectos es avasallador. Se puede aplicar para hallar la probabilidad de cualquier evento definido para un espacio muestral que contenga un conjunto finito o contable de puntos muestrales, pero no es resistente al error humano. Los errores comunes incluyen diagnosticar de manera incorrecta la naturaleza de un evento simple y no indicar todos los puntos muestrales en S. Una segunda complicación se presenta porque muchos espacios muestrales contienen un gran número de puntos muestrales y un detallado completo es tedioso y lento y podría ser prácticamente imposible.

Por fortuna, numerosos espacios muestrales generados por datos experimentales contienen subconjuntos de puntos muestrales que son igualmente probables. (Los espacios muestrales para los Ejemplos 2.1, 2.2 y 2.3 poseen esta propiedad.) Cuando esto ocurre, no es necesario indicar los puntos pero podemos contar el número de cada subconjunto. Si no se pueden aplicar estos métodos de conteo, debe usarse un método ordenado para indicar los puntos muestrales (obsérvense los esquemas de lista para los Ejemplos 2.1, 2.2 y 2.3). Enlistan números grandes de puntos muestrales se puede lograr con el uso de una computadora.

Las herramientas que reducen el esfuerzo y error asociado con el método de punto muestral para hallar la probabilidad de un evento incluyen el sentido de orden, una computadora y la teoría matemática de conteo llamada *análisis combinatorio.* La programación y aplicaciones en computadora forman un tema de estudio por separado. La teoría matemática de análisis combinatorio también es un tema muy amplio, pero se pueden obtener resultados bastante útiles en forma breve. En consecuencia, nuestro siguiente tema se refiere a algunos resultados elementales en análisis combinatorio y su aplicación al método de punto muestral para la solución de problemas de probabilidad.

Ejercicios

2.25 Al azar se selecciona un solo auto de entre todos los registrados en una agencia local de placas. ¿Qué piensa usted de la siguiente frase? "Todos los autos son marca Volkswagen o no lo son. Por tanto, la probabilidad de que el auto seleccionado sea un Volkswagen es de 1/2."

2.26 Según el *Webster's New Collegiate Dictionary*, una varilla adivinadora es "una barra en forma de horqueta que se piensa indica (o adivina) la presencia de agua o minerales al inclinarse hacia abajo cuando se mantiene sobre una veta". Para probar el dicho de un experto en varillas adivinadoras, unos escépticos entierran cuatro latas en el suelo, dos vacías y dos llenas de agua. El experto es llevado a las cuatro latas y se le dice que dos de ellas contienen agua. Él usa la varilla adivinadora para probar cada una de las cuatro latas y decide cuáles de ellas contienen agua.

a Indique el espacio muestral para este experimento.

b Si la varilla adivinadora es totalmente inútil para localizar agua, ¿cuál es la probabilidad de que el experto identifique en forma correcta (por adivinación) las dos latas que contienen agua?

2.27 En el Ejercicio 2.12 consideramos una situación en la que unos autos entran a un crucero y podría cada uno de ellos dar vuelta a la derecha, a la izquierda o seguir de frente. Un experimento consiste en observar dos vehículos que pasan por el crucero.

a ¿Cuántos puntos muestrales hay en el espacio muestral? Indíquelos.

b Suponiendo que todos los puntos muestrales sean igualmente probables, ¿cuál es la probabilidad de que al menos un auto dé vuelta a la izquierda?

c De nuevo suponiendo puntos muestrales igualmente probables, ¿cuál es la probabilidad de que a lo sumo un vehículo dé vuelta?

2.28 Cuatro personas igualmente calificadas hacen solicitud para ocupar dos puestos idénticos en una empresa. Un y sólo un solicitante es miembro de un grupo de minoría étnica. Los puestos se llenan al seleccionar dos de los solicitantes al azar.

a Indique los posibles resultados para este experimento.

b Asigne probabilidades razonables a los puntos muestrales.

c Encuentre la probabilidad de que el solicitante del grupo étnico minoritario sea seleccionado para una posición.

2.29 Se necesitan dos jurados adicionales para completar un jurado para un juicio criminal. Hay seis jurados en perspectiva, dos mujeres y cuatro hombres. Dos de los jurados son seleccionados al azar de entre los seis disponibles.

a Defina el experimento y describa un punto muestral. Suponga que es necesario describir sólo los dos jurados seleccionados y no el orden en el que fueron elejidos.

b Indique el espacio muestral asociado con este experimento.

c ¿Cuál es la probabilidad de que los dos jurados seleccionados sean mujeres?

2.30 Tres vinos importados van a ser clasificados de menos a más por un experto en vinos. Esto es, un vino será identificado como el mejor, otro como el segundo mejor y el vino restante como el peor.

a Describa un punto muestral para este experimento.

b Indique el espacio muestral.

c Suponga que el "experto" no sabe en realidad nada de vinos y que al azar asigna el lugar a los tres vinos. Uno de los vinos es de mucha mejor calidad que los otros. ¿Cuál es la probabilidad de que el experto no clasifique el mejor vino como peor que el segundo mejor?

2.31 Un furgón de ferrocarril contiene seis sistemas electrónicos complejos. Dos de los seis se han de seleccionar al azar para hacerles pruebas completas y luego clasificarlos como defectuosos o no defectuosos.

a Si dos de los seis sistemas en realidad están defectuosos, encuentre la probabilidad de que al menos uno de los dos sistemas probados sea defectuoso. Encuentre la probabilidad de que ambos sean defectuosos.

b Si cuatro de los seis sistemas están defectuosos en realidad, encuentre las probabilidades indicadas en el inciso a.

2.32 Un detallista vende sólo dos estilos de consolas estéreo y la experiencia muestra que ambas tienen igual demanda. Cuatro clientes en sucesión entran a la tienda para comprar estéreos. El vendedor está interesado en sus preferencias.

a Indique las posibilidades para arreglos de preferencia entre los cuatro clientes (esto es, indique el espacio muestral).

b Asigne probabilidades a los puntos muestrales.

c Denote con A el evento de que los cuatro clientes prefieran el mismo estilo. Encuentre $P(A)$.

2.33 La Oficina del Censo reporta que el ingreso familiar mediano para todas las familias en Estados Unidos, durante el año 2003, fue \$43,318. Esto es, la mitad de todas las familias estadounidenses tuvo ingresos que rebasaban esta cantidad y la mitad tuvo ingresos iguales o menores a esta cantidad. Suponga que se hace una encuesta a cuatro familias y que cada una de ellas revela si su ingreso rebasó los \$43,318 en 2003.

a Indique los puntos del espacio muestral.

b Identifique los eventos simples en cada uno de los eventos siguientes:

A: al menos dos tuvieron ingresos de más de \$43,318.

B: exactamente dos tuvieron ingresos de más de \$43,318.

C: exactamente una tuvo ingresos menores o iguales a \$43,318.

c Haga uso de la interpretación dada para la mediana para asignar probabilidades a los eventos simples y encuentre $P(A)$, $P(B)$ y $P(C)$.

2.34 Los pacientes que llegan a una clínica para atención externa pueden seleccionar una de tres estaciones de servicio. Suponga que los médicos se asignan al azar a las estaciones y que los pacientes por tanto no tienen preferencia de estación. Tres pacientes llegan a la clínica y se observa su selección de estación.

a Indique los puntos muestrales para el experimento.

b Sea A el evento de que cada estación recibe a un paciente. Indique los puntos muestrales en A.

c Haga una asignación razonable de probabilidades a los puntos muestrales y encuentre $P(A)$.

2.6 Herramientas para contar puntos muestrales

Esta sección presenta algunos resultados útiles de la teoría de análisis combinatorio e ilustra la aplicación de ellos al método de punto muestral para hallar la probabilidad de un evento. En muchos casos estos resultados hacen posible contar el número total de puntos muestrales en el espacio muestral S y en un evento de interés, con lo cual dan una confirmación de la lista de eventos simples. Cuando el número de eventos simples de un espacio muestral es muy grande y la enumeración manual de todo punto muestral es tediosa o hasta imposible, contar el número de puntos del espacio muestral y del evento de interés puede ser la única forma eficiente de calcular la probabilidad de un evento. De hecho, si un espacio muestral contiene N puntos muestrales igualmente probables y un evento A contiene exactamente n_a puntos muestrales, es fácil ver que $P(A) = n_a/N$.

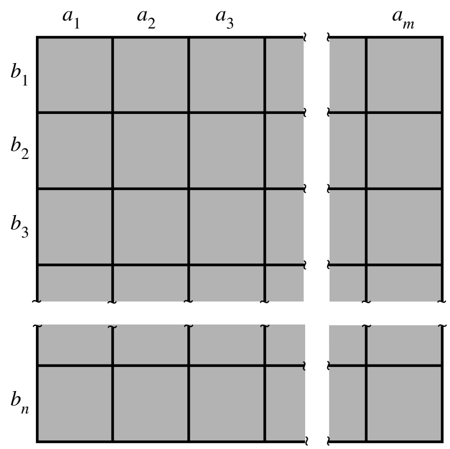


FIGURA **2.9** Tabla que indica el número de pares (a_i, b_j)

El primer resultado del análisis combinatorio que presentamos, a veces llamado *regla mn*, se expresa como sigue:

TEOREMA **2.1** Con m elementos $a_1, a_2, \ldots, a_m$ y n elementos $b_1, b_2, \ldots, b_n$, es posible formar $mn = m \times n$ pares que contengan un elemento de cada grupo.

Demostración La verificación del teorema se puede ver al observar la tabla rectangular de la Figura 2.9. Hay un cuadro de la tabla para cada par a_i, b_j y por tanto un total de $m \times n$ cuadros.

La regla *mn* puede ser extendida a cualquier número de conjuntos. Dados tres conjuntos de elementos $a_1, a_2, \ldots, a_m$; $b_1, b_2, \ldots, b_n$; y $c_1, c_2, \ldots, c_p$ el número de ternas distintas que contienen un elemento de cada conjunto es igual a *mnp*. La prueba del teorema para tres conjuntos invulucra dos aplicaciones del Teorema 2.1. Consideramos el primer conjunto como un par (a_i, b_j) y unimos cada uno de estos pares con elementos del tercer conjunto, $c_1, c_2, \ldots, c_p$. El Teorema 2.1 implica que hay *mn* pares (a_i, b_j). Como hay p elementos $c_1, c_2, \ldots, c_p$, otra aplicación del Teorema 2.1 implica que hay $(mn)(p) = mnp$ ternas $a_i b_j c_k$.

EJEMPLO **2.5** Un experimento incluye lanzar un par de dados y observar los números de sus caras superiores. Encuentre el número de puntos muestrales en S, el espacio muestral para el experimento.

Solución Un punto muestral para este experimento puede ser representado simbólicamente como un par ordenado de números que representan los resultados en el primero y segundo dados, respectivamente. Así, (4, 5) denota el evento de que la cara superior del primer dado fue un 4 y en el segundo dado, un 5. El espacio muestral S está formado por el conjunto de todos los pares posibles (x, y), donde x y y son ambos enteros entre 1 y 6.

El primer dado puede resultar en uno de seis números. Éstos representan $a_1, a_2, \ldots, a_6$. De igual modo, el segundo dado puede caer en una de seis formas y éstas corresponden a $b_1, b_2, \ldots, b_6$. Entonces $m = n = 6$ y el número total de puntos muestrales en S es $mn = (6)(6) = 36$. ■

EJEMPLO **2.6** Consulte el experimento de lanzar al aire una moneda del Ejemplo 2.3. Encontramos para este ejemplo que el número total de puntos muestrales era ocho. Use la extensión de la regla *mn* para confirmar este resultado.

Solución Cada punto muestral en S fue identificado por una secuencia de tres letras, cada posición de la secuencia contenía una de dos letras, una H o una T. El problema entonces implica la formación de ternas, con un elemento (una H o una T) de cada uno de los tres conjuntos. Para este ejemplo, los conjuntos son idénticos y todos contienen dos elementos (H y T). Así, el número de elementos en cada conjunto es $m = n = p = 2$, y el número total de ternas que se puede formar es $mnp = (2)^3 = 8$. ■

EJEMPLO **2.7** Considere un experimento que consiste en registrar el cumpleaños para cada una de 20 personas seleccionadas al azar. Si no se presta atención a los años bisiestos y se supone que hay sólo 365 cumpleaños distintos posibles, encuentre el número de puntos del espacio muestral S para este experimento. Si suponemos que cada uno de los posibles conjuntos de cumpleaños es igualmente probable, ¿cuál es la probabilidad de que cada persona de las 20 tenga un cumpleaños diferente?

Solución Numere los días del año como $1, 2, \ldots, 365$. Un punto muestral para este experimento puede ser representado por una sucesión ordenada de 20 números, donde el primer número denota el número del día que es el cumpleaños de la primera persona, el segundo número implica el número del día que es el cumpleaños de la segunda persona, y así sucesivamente. Estamos interesados en el número de veintenas que se pueden formar, seleccionando un número que represente uno de los 365 días del año para cada uno de los 20 conjuntos. Los conjuntos son todos idénticos y cada uno contiene 365 elementos. Aplicaciones repetidas de la regla *mn* nos dice que hay $(365)^{20}$ de tales veintenas. Entonces, el espacio muestral S contiene $N = (365)^{20}$ puntos muestrales. Aun cuando no podríamos de manera factible hacer una lista de todos los puntos muestrales, sí suponemos que son igualmente probables, $P(E_i) = 1/(365)^{20}$ para cada evento simple.

Si denotamos por A el evento de que cada persona tenga un cumpleaños diferente, la probabilidad de A se puede calcular si podemos determinar n_a, el número de puntos muestrales en A. Un punto muestral está en A si la correspondiente veintena es tal que no hay dos posiciones que contengan el mismo número. Entonces, el conjunto de números del cual se puede seleccionar el primer elemento en la veintena de A contiene 365 números, el conjunto del cual se puede seleccionar el segundo elemento contiene 364 elementos (todos excepto el seleccionado para el primer elemento), el conjunto del cual se puede seleccionar el tercero contiene 363 (todos excepto los dos seleccionados para los primeros dos elementos), … y el conjunto del cual se puede seleccionar el vigésimo elemento contiene 346 elementos (todos excepto los seleccionados para los primeros 19 elementos). Una extensión de la regla *mn* dará

$$n_a = (365) \times (364) \times \cdots \times (346).$$

Por último, podemos determinar que

$$P(A) = \frac{n_a}{N} = \frac{365 \times 364 \times \cdots \times 346}{(365)^{20}} = .5886.$$

■

Observe que para los Ejemplos 2.5 y 2.6 los números de puntos muestrales en los respectivos espacios muestrales son ambos relativamente pequeños y que las listas para estos espacios muestrales podrían escribirse con facilidad. Para ejemplos como éstos, la regla *mn* proporciona un método simple para verificar que los espacios muestrales contienen el número correcto de puntos. En contraste, no es factible hacer una lista del espacio muestral en el Ejemplo 2.7. No obstante, la regla *mn* se puede usar para contar el número de puntos muestrales en S y en el evento de interés, permitiendo el cálculo de la probabilidad del evento.

Hemos visto que los puntos muestrales asociados con un experimento a veces pueden representarse de manera simbólica como una sucesión de números o símbolos. En algunos casos será evidente que el número total de puntos muestrales es igual al número de formas distintas en que los respectivos símbolos se pueden acomodar en sucesión. El siguiente teorema se puede usar para determinar el número de arreglos ordenados que se pueden formar.

DEFINICIÓN **2.7**

Un arreglo ordenado de r objetos distintos se denomina *permutación.* El número de formas de ordenar n objetos distintos tomados r a la vez estará designado por el símbolo P_r^n.

TEOREMA **2.2**

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Demostración

Estamos interesados en el número de formas de llenar r posiciones con n objetos distintos. Aplicando la extensión de la regla *mn*, vemos que el primer objeto se puede seleccionar en una de n formas. Después de escoger el primero, el segundo se puede escoger en $(n-1)$ formas, el tercero en $(n-2)$ y el r-ésimo en $(n-r+1)$ formas. En consecuencia, el número total de arreglos distintos es

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1).$$

Expresado en términos de factoriales,

$$P_r^n = n(n-1)(n-2)\cdots(n-r+1)\frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

donde $n! = n(n-1)\cdots(2)(1)$ y $0! = 1$.

EJEMPLO **2.8** Los nombres de 3 empleados se han de sacar al azar, sin restitución, de un tazón que contiene los nombres de 30 empleados de una pequeña compañía. La persona cuyo nombre sea sacado primero recibe \$100 y aquellos cuyos nombres se saquen en segundo y tercero recibirán \$50 y \$25, respectivamente. ¿Cuántos puntos muestrales están asociados con este experimento?

Solución Debido a que los premios otorgados son diferentes, el número de puntos muestrales es el número de arreglos ordenados de $r = 3$ de entre los posibles $n = 30$ nombres. Entonces, el número de puntos muestrales en S es

$$P_3^{30} = \frac{30!}{27!} = (30)(29)(28) = 24{,}360.$$

■

EJEMPLO **2.9** Suponga que una operación de ensamble en una planta de manufacturas consta de cuatro pasos que se pueden efectuar en cualquier secuencia. Si el fabricante desea comparar el tiempo de ensamble para cada una de las secuencias, ¿cuántas secuencias diferentes estarán involucradas en el experimento?

Solución El número total de secuencias es igual al número de formas de acomodar los $n = 4$ pasos tomados $r = 4$ a la vez, o sea

$$P_4^4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = 24. \qquad \blacksquare$$

El siguiente resultado de análisis combinatorio se puede usar para determinar el número de subconjuntos de varios tamaños que se pueden formar al dividir un conjunto de n objetos distintos en k grupos que no se traslapen.

TEOREMA **2.3** El número de formas de dividir n objetos distintos en k grupos distintos que contienen $n_1, n_2, \ldots, n_k$ objetos, respectivamente, donde cada objeto aparece en exactamente un grupo y $\sum_{i=1}^{k} n_i = n$, es

$$N = \binom{n}{n_1\ n_2\ \cdots\ n_k} = \frac{n!}{n_1!\, n_2!\, \cdots\, n_k!}.$$

Demostración N es el número de arreglos distintos de n objetos en una fila para un caso en el que el reacomodo de los objetos dentro de un grupo no cuenta. Por ejemplo, las letras de la a a la l están acomodadas en tres grupos, donde $n_1 = 3$, $n_2 = 4$ y $n_3 = 5$:

$$abc|de|f|g|hijkl$$

es uno de estos arreglos.

El número de arreglos distintos de los n objetos, suponiendo que todos los objetos sean distintos, es $P_n^n = n!$ (del Teorema 2.2). Entonces P_n^n es igual al número de formas de dividir los n objetos en k grupos (ignorando el orden dentro de los grupos) multiplicado por el número de formas de ordenar los $n_1, n_2, \ldots, n_k$ elementos dentro de cada grupo. Esta aplicación de la regla *mn* extendida da

$$P_n^n = (N) \cdot (n_1!\, n_2!\, n_3! \cdots n_k!),$$

donde $n_i!$ es el número de arreglos distintos de los n_i objetos del grupo i.

Al despejar N, tenemos

$$N = \frac{n!}{n_1!\, n_2! \cdots n_k!} \equiv \binom{n}{n_1\ n_2\ \cdots\ n_k}.$$

Los términos $\binom{n}{n_1\ n_2 \cdots n_k}$ se denominan *coeficientes multinomiales* porque se presentan en la expansión del *término multinomial* $y_1 + y_2 + \cdots + y_k$ elevada a la n potencia:

$$(y_1 + y_2 + \cdots + y_k)^n = \sum \binom{n}{n_1\ n_2\ \cdots\ n_k} y_1^{n_1} y_2^{n_2} \cdots y_k^{n_k},$$

donde esta suma se toma sobre toda $n_i = 0, 1, \ldots, n$ tal que $n_1 + n_2 + \cdots + n_k = n$.

EJEMPLO **2.10** Ha surgido una disputa laboral respecto a la distribución de 20 trabajadores a cuatro trabajos de construcción diferentes. El primer trabajo (considerado muy indeseable) requirió de 6 trabajadores; el segundo, tercero y cuarto utilizaron 4, 5 y 5 trabajadores, respectivamente. La disputa surgió sobre una supuesta distribución aleatoria de los trabajadores a los trabajos que pusieron a los 4 miembros de un grupo étnico particular en el trabajo 1. Al considerar si la asignación representaba una injusticia, un panel de mediación pedía la probabilidad del evento observado. Determine el número de puntos muestrales del espacio muestral S para este experimento. Esto es, determine el número de formas en que los 20 trabajadores se pueden dividir en grupos de los tamaños apropiados para llenar todas las posiciones de trabajo. Encuentre la probabilidad del evento observado si se supone que los trabajadores son asignados en forma aleatoria a los trabajos.

Solución El número de formas de asignar los 20 trabajadores a los cuatro trabajos es igual al número de formas de dividir los 20 en cuatro grupos de tamaños $n_1 = 6$, $n_2 = 4$, $n_3 = n_4 = 5$. Entonces

$$N = \binom{20}{6\ 4\ 5\ 5} = \frac{20!}{6!\,4!\,5!\,5!}.$$

Por una *asignación aleatoria* de trabajadores a los trabajos queremos decir que cada uno de los N puntos muestrales tiene probabilidad igual a $1/N$. Si A denota el evento de interés y n_a el número de puntos muestrales en A, la suma de las probabilidades de los puntos muestrales en A es $P(A) = n_a(1/N) = n_a/N$. El número de puntos muestrales en A, n_a, es el número de formas de asignar trabajadores a los cuatro trabajos, con los 4 miembros del grupo étnico pasando todos al trabajo 1. Los 16 trabajadores restantes necesitan ser asignados a los trabajos restantes. Debido a que quedan dos vacantes para el trabajo 1, esto se puede hacer en

$$n_a = \binom{16}{2\ 4\ 5\ 5} = \frac{16!}{2!\,4!\,5!\,5!}$$

formas. Se deduce que

$$P(A) = \frac{n_a}{N} = 0.0031.$$

Entonces, si los trabajadores se asignan al azar a los trabajos, la probabilidad de que los 4 miembros del grupo étnico pasen al trabajo indeseable es muy pequeña. Hay razones para dudar de que los trabajos se asignaron al azar. ■

En muchas situaciones los puntos muestrales son identificados por un conjunto de símbolos en los que el arreglo de símbolos *no es importante*. Los puntos muestrales para la selección de solicitantes, Ejemplo 2.2, implica una selección de dos solicitantes de entre cinco. Cada punto muestral es identificado como un par de símbolos y el orden de los símbolos empleados para identificar los puntos muestrales no es importante.

DEFINICIÓN **2.8**

El número de *combinaciones* de n objetos tomados r a la vez es el número de subconjuntos, cada uno de tamaño r, que se pueden formar a partir de los n objetos. Este número estará denotado por C_r^n o $\binom{n}{r}$.

TEOREMA **2.4**

El número de subconjuntos desordenados de tamaño r escogidos (sin restitución) de n objetos disponibles es

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}.$$

Demostración

La selección de r objetos de un total de n objetos es equivalente a dividir los n objetos en $k = 2$ grupos, los r seleccionados y los $(n - r)$ restantes. Éste es el caso especial del problema general de división tratado con el Teorema 2.3. En el presente caso, $k = 2$, $n_1 = r$ y $n_2 = (n - r)$ y, por tanto,

$$\binom{n}{r} = C_r^n = \binom{n}{r \quad n-r} = \frac{n!}{r!(n-r)!}.$$

Los términos $\binom{n}{r}$ se conocen generalmente como *coeficientes binomiales* porque se presentan en la *expansión binomial*

$$\begin{aligned}(x+y)^n &= \binom{n}{0}x^n y^0 + \binom{n}{1}x^{n-1}y^1 + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}x^0 y^n \\ &= \sum_{i=0}^{n}\binom{n}{i}x^{n-i}y^i.\end{aligned}$$

EJEMPLO **2.11** Encuentre el número de formas de seleccionar dos solicitantes de entre cinco y por tanto el número total de puntos muestrales en S para el Ejemplo 2.2.

Solución

$$\binom{5}{2} = \frac{5!}{2!3!} = 10.$$

(Observe que éste está acorde con el número de puntos muestrales indicados en el Ejemplo 2.2.) ■

EJEMPLO **2.12** Denote con A el evento de que exactamente uno de los dos mejores solicitantes aparezca en una selección de dos de entre cinco. Encuentre el número de puntos muestrales en A y $P(A)$.

Solución Denote con n_a el número de puntos muestrales en A. Entonces n_a es igual al número de formas de seleccionar uno de los dos mejores (llame m a este número) multiplicado por el número de formas de seleccionar uno de los tres solicitantes de baja calificación (llame n a este número). Entonces $m = \binom{2}{1}$, $n = \binom{3}{1}$ y, aplicando la regla mn,

$$n_a = \binom{2}{1} \cdot \binom{3}{1} = \frac{2!}{1!1!} \cdot \frac{3!}{1!2!} = 6.$$

(Este número se puede verificar al contar los puntos muestrales en A de la lista del Ejemplo 2.2.)

En el Ejemplo 2.11 hallamos que el número total de puntos muestrales en S es $N = 10$. Si cada selección es igualmente probable, $P(E_i) = 1/10 = .1$, $i = 1, 2, \ldots, 10$ y

$$P(A) = \sum_{E_i \subset A} P(E_i) = \sum_{E_i \subset A} (.1) = n_a(.1) = 6(.1) = .6.$$ ■

EJEMPLO **2.13** Una empresa compra abastecimientos a M distribuidores y desea colocar n pedidos ($n < M$). Suponga que la empresa coloca los pedidos en forma que permita a cada distribuidor tener igual probabilidad de obtener cualquier pedido y no hay restricción en el número de pedidos que se puedan colocar con cualquier distribuidor. Encuentre la probabilidad de que un distribuidor particular, por ejemplo el distribuidor I, obtenga exactamente k pedidos ($k \leq n$).

Solución Como cualquiera de los M distribuidores puede ser seleccionado para recibir cualquiera de los pedidos, hay M formas en que cada pedido se pueda colocar y el número de formas diferentes en que los n pedidos se pueden colocar es $M \cdot M \cdot M \cdots = (M)^n$. En consecuencia, hay $(M)^n$ puntos muestrales en S. Todos estos puntos son igualmente probables; en consecuencia, $P(E_i) = 1/(M)^n$.

Denote con A el evento de que el distribuidor I reciba exactamente k pedidos de entre los n. Los k pedidos asignados al distribuidor I se pueden seleccionar de los n en $\binom{n}{k}$ formas. Resta determinar el número de formas en que los restantes $(n - k)$ pedidos se puedan asignar a los otros $M - 1$ distribuidores. Como cada uno de estos $(n - k)$ pedidos pueden ir a cualquiera de los $(M - 1)$ distribuidores, esta asignación se puede hacer en $(M - 1)^{n-k}$ formas. Entonces, A contiene

$$n_a = \binom{n}{k}(M - 1)^{n-k}$$

puntos muestrales, y como los puntos muestrales son igualmente probables,

$$P(A) = \sum_{E_i \subset A} P(E_i) = \sum_{E_i \subset A} \left(\frac{1}{M^n}\right) = n_a \left(\frac{1}{M^n}\right) = \frac{\binom{n}{k}(M - 1)^{n-k}}{M^n}.$$ ■

Los Teoremas 2.1 a 2.4 dan algunas de las numerosas y útiles formas de conteo halladas en la teoría de análisis combinatorio. Unos cuantos teoremas adicionales aparecen en los ejercicios al final de este capítulo. Si el lector está interesado en ampliar su conocimiento de análisis combinatorio, consulte uno de los numerosos textos sobre este tema.

A continuación dirigiremos nuestra atención al concepto de probabilidad condicional. La probabilidad condicional desempeña un importante papel en el método de composición de eventos para hallar la probabilidad de un evento y en ocasiones es útil para hallar las probabilidades de puntos muestrales (para espacios muestrales con puntos muestrales que no son igualmente probables).

Ejercicios

2.35 Una línea aérea tiene seis vuelos de Nueva York a California y siete vuelos de California a Hawai diarios. Si los vuelos se han de hacer en días separados, ¿cuántos arreglos diferentes de vuelos puede ofrecer la línea aérea de Nueva York a Hawai?

2.36 Una operación de ensamble en una planta de manufactura requiere tres pasos que se pueden realizar en cualquier secuencia. ¿En cuántas formas diferentes se puede efectuar el ensamble?

2.37 Una mujer de negocios de Filadelfia está preparando su itinerario para una visita a seis ciudades importantes. La distancia recorrida y por tanto el costo del viaje, dependerá del orden en que ella planee su ruta.

a ¿Cuántos itinerarios diferentes (y costos de viaje) son posibles?

b Si la mujer selecciona al azar uno de los posibles itinerarios y Denver y San Francisco son dos de las ciudades que ella piensa visitar, ¿cuál es la probabilidad de que visite Denver antes de San Francisco?

2.38 Un restaurante de nivel económico alto ofrece un menú especial de *precios fijos* en el que, por un costo fijo de comidas, una persona puede seleccionar de entre cuatro aperitivos, tres ensaladas, cuatro entradas y cinco postres. ¿Cuántas comidas diferentes hay si una de ellas consta de un aperitivo, una ensalada, una entrada y un postre?

2.39 Un experimento consiste en tirar un par de dados.

a Use los teoremas combinatorios para determinar el número de puntos muestrales del espacio muestral S.

b Encuentre la probabilidad de que la suma de los números que aparezcan en el dado sea igual a 7.

2.40 Una marca de automóvil viene en cinco estilos diferentes, con cuatro tipos de motor, con dos tipos de transmisiones y en ocho colores.

a ¿Cuántos autos tendría que tener en existencia un distribuidor si incluyera uno por cada combinación de estilo, motor y transmisión?

b ¿Cuántos tendría que tener en existencia un centro de distribución si todos los colores de autos se tuvieran para cada combinación del inciso a?

2.41 ¿Cuántos números telefónicos diferentes de siete dígitos se pueden formar si el primer dígito no puede ser cero?

2.42 La directora de personal de una corporación ha contratado diez nuevos ingenieros. Si tres puestos de trabajo (muy distintos) se abren en una planta en Cleveland, ¿en cuántas formas puede ella ocupar los puertos?

2.43 Una flota de nueve taxis se ha de despachar a tres aeropuertos en forma tal que tres vayan al aeropuerto A, cinco al aeropuerto B y uno al aeropuerto C. ¿En cuántas formas distintas se puede lograr esto?

2.44 Consulte el Ejercicio 2.43. Suponga que los taxis son asignados a aeropuertos al azar.

a Si exactamente uno de los taxis necesita reparación, ¿cuál es la probabilidad de que sea despachado al aeropuerto C?

b Si exactamente tres de los taxis necesitan reparación, ¿cuál es la probabilidad de que cada aeropuerto reciba uno de los taxis que necesita reparación?

2.45 Suponga que deseamos expandir $(x + y + z)^{17}$. ¿Cuál es el coeficiente de $x^2y^5z^{10}$?

2.46 Diez equipos están jugando en un torneo de baloncesto. En la primera ronda, a los equipos se les asignan al azar los juegos 1, 2, 3, 4 y 5. ¿En cuántas formas pueden ser asignados los equipos a los juegos?

***2.47** Consulte el Ejercicio 2.46. Si $2n$ equipos van a ser asignados a los juegos $1, 2, \ldots, n$, ¿en cuántas formas pueden ser asignados los equipos a los juegos?

2.48 Si deseamos expandir $(x + y)^8$, ¿cuál es el coeficiente de x^5y^3? ¿Cuál es el coeficiente de x^3y^5?

2.49 Los estudiantes que asisten a clase en la Universidad de Florida pueden seleccionar de entre 130 disciplinas de posgrado. El posgrado de un estudiante se identifica en los registros del secretario general de la universidad con un código de dos o tres letras (por ejemplo, los posgrados en estadística se identifican con STA, los de matemáticas con MS). Algunos estudiantes optan por dos posgrados y completan los requisitos para ambos antes de graduarse. Al secretario general se le pidió considere asignar a estos dos posgrados un código diferente de dos o tres letras para que puedan ser identificados en el sistema de registro del estudiante.

a ¿Cuál es el número máximo de posibles posgrados dobles disponibles para estudiantes de la Universidad de Florida?

b Si existe cualquier código de dos o tres letras para identificar posgrados individuales o dobles, ¿de cuántos códigos de posgrado se dispone?

c ¿Cuántos códigos de posgrado se requieren para identificar estudiantes que tienen ya sea un posgrado individual o uno doble?

d ¿Hay suficientes códigos de posgrado para identificar todos los posgrados individuales o dobles en la Universidad de Florida?

2.50 La probabilidad desempeñó un papel en la manipulación de la lotería del estado de Pennsylvania del 24 de abril de 1980 (*Los Angeles Times*, 8 de septiembre de 1980). Para determinar cada uno de los dígitos del número ganador de tres dígitos, cada uno de los números 0, 1, 2,…, 9 se coloca en una pelota de ping-pong, las diez pelotas se agitan con aire en un compartimento y el número seleccionado para el dígito es el de la bola que flota a la parte superior de la máquina. Para alterar las probabilidades, los conspiradores inyectaron un líquido en todas las bolas empleadas en el juego excepto las que tenían los números 4 y 6, haciendo casi seguro que las bolas más livianas fueran seleccionadas y determinar los dígitos del número ganador. Luego compraron billetes de lotería con los potenciales números ganadores. ¿Cuántos números ganadores potenciales hubo (al final 666 fue el ganador)?

2.51 Una fraternidad local está realizando una rifa en la que se han de vender 50 boletos, uno por cliente. Hay tres premios para ser concedidos. Si los cuatro organizadores de la rifa compran un boleto cada uno, ¿cuál es la probabilidad de que los cuatro organizadores ganen

a todos los premios?,

b exactamente dos de los premios?,

c exactamente uno de los premios?,

d ninguno de los premios?

2.52 Un experimentador desea investigar el efecto de tres variables: presión, temperatura y tipo de catalizador en el rendimiento de un proceso de refinación. Si el experimentador trata de usar tres ajustes, cada uno para temperatura y presión y dos tipos de catalizadores, ¿cuántas series experimentales tendrán que ejecutarse si él desea realizar todas las posibles combinaciones de presión, temperatura y tipos de catalizadores?

2.53 Cinco compañías, $F_1, F_2, \ldots, F_5$, ofrecen cotizaciones para tres contratos separados C_1, C_2 y C_3. A cualquiera de las firmas se le otorgará un contrato a lo sumo. Los contratos son muy diferentes, de modo que una asignación de C_1 a F_1, por ejemplo, debe ser distinta de una asignación de C_2 a F_1.

a ¿Cuántos puntos muestrales hay en total en este experimento que involucra la asignación de contratos a las firmas? (No es necesario indicarlos todos.)

b Dada la suposición de puntos muestrales igualmente probables, encuentre la probabilidad de que a F_3 se le conceda un contrato.

2.54 Hay un grupo de tres pasantes y cinco estudiantes egresados para ocupar ciertos puestos gubernamentales para estudiantes. Si cuatro estudiantes se han de seleccionar al azar de entre este grupo, encuentre la probabilidad de que exactamente dos pasantes se encuentren entre los cuatro seleccionados.

2.55 Se ha de realizar un estudio en un hospital para determinar las actitudes de los enfermeros hacia diversos procedimientos administrativos. Se ha de seleccionar una muestra de 10 enfermeros de entre un total de 90 enfermeros empleadas por el hospital.

a ¿Cuántas muestras diferentes de 10 enfermeros se pueden seleccionar?

b Veinte de los 90 enfermeros son hombres. Si 10 enfermeros se seleccionan al azar entre los empleados por el hospital, ¿cuál es la probabilidad de que la muestra de diez incluirá exactamente 4 hombres (y 6 mujeres) enfermeros?

2.56 Una estudiante se prepara para un examen estudiando una lista de diez problemas. Ella puede resolver seis de ellos. Para el examen, el profesor selecciona cinco problemas al azar de los diez de la lista dada a los estudiantes. ¿Cuál es la probabilidad de que la estudiante pueda resolver los cinco problemas del examen?

2.57 Se sacan dos cartas de una baraja estándar de 52 cartas. ¿Cuál es la probabilidad de obtener un as y una figura?

2.58 Se reparten cinco cartas de una baraja de 52 cartas. ¿Cuál es la probabilidad de sacar

a 3 ases y 2 reyes?,

b un "full" (tres cartas de una clase, 2 cartas de otra clase)?

2.59 Se reparten cinco cartas de una baraja de 52 cartas. ¿Cuál es la probabilidad de sacar

a 1 as, 1 dos, 1 tres, 1 cuatro y 1 cinco (esta es una forma de obtener una "escalera")?,

b cualquier escalera?

2.60 Consulte el Ejemplo 2.7. Suponga que registramos el nacimiento para cada una de las n personas seleccionadas al azar.

a Escriba una expresión para la probabilidad de que ninguna comparta el mismo cumpleaños.

b ¿Cuál es el valor más pequeño de n para que la probabilidad sea al menos .5 de que como mínimo dos personas comparten el mismo cumpleaños?

2.61 Suponga que preguntamos a n personas seleccionadas al azar si cumplen años el mismo día que usted.

a Escriba una expresión para la probabilidad de que nadie comparta su cumpleaños con usted (pase por alto los años bisiestos).

b ¿Cuántas personas es necesario seleccionar para que la probabilidad sea al menos .5 de que como mínimo una comparta cumpleaños con usted?

2.62 Un fabricante tiene nueve motores distintos en existencia, dos de los cuales llegaron de un proveedor en particular. Los motores deben dividirse entre tres líneas de producción, con tres motores pasando a cada línea. Si la asignación de motores a líneas es aleatoria, encuentre la probabilidad de que ambos motores del proveedor particular sean asignados a la primera línea.

2.63 Los ocho miembros del Consejo Asesor de Relaciones Humanas de Gainesville, Florida, consideró la queja de una mujer que alegaba discriminación, basada en su género, de parte de una empresa local. El Consejo, compuesto de cinco mujeres y tres hombres, votó 5–3 a favor de la quejosa, con las cinco mujeres votando a favor de ella y los tres hombres en contra. El abogado que representaba a la compañía apeló la decisión del Consejo reclamando sesgo de género de parte de los miembros del Consejo. Si no hubo sesgo de género entre los miembros del Consejo, podría ser razonable hacer conjeturas de que sería probable que cualquier grupo de cinco miembros votara a favor de la quejosa como lo haría cualquier otro grupo de cinco. Si éste fuera el caso, ¿cuál es la probabilidad de que el voto se dividiera por líneas de género (cinco mujeres a favor, tres hombres en contra)?

2.64 Un dado balanceado se tira seis veces y cada vez se registra el número de su cara superior, ¿Cuál es la probabilidad de que los números registrados sean 1, 2, 3, 4, 5 y 6 en cualquier orden?

2.65 Consulte el Ejercicio 2.64. Suponga que el dado ha sido alterado para que las caras sean 1, 2, 3, 4, 5 y 5. Si el dado se tira cinco veces, ¿cuál es la probabilidad de que los números registrados sean 1, 2, 3, 4 y 5 en cualquier orden?

2.66 Consulte el Ejemplo 2.10. ¿Cuál es la probabilidad de que

a el miembro de un grupo étnico sea asignado a cada tipo de trabajo?,

b no se asigne ningún miembro de un grupo étnico a un trabajo tipo 4?

2.67 Consulte el Ejemplo 2.13. Suponga que el número de distribuidores es $M = 10$ y que hay $n = 7$ pedidos por ser colocados. ¿Cuál es la probabilidad de que

a todos los pedidos vayan a distribuidores diferentes?,

***b** el distribuidor I obtenga exactamente dos pedidos *y* el distribuidor II obtenga exactamente tres pedidos?,

***c** los distribuidores I, II y III obtengan exactamente dos, tres y un pedido(s), respectivamente?

2.68 Demuestre que, para cualquier entero $n \geq 1$,

a $\binom{n}{n} = 1$. Interprete este resultado.

b $\binom{n}{0} = 1$. Interprete este resultado.

c $\binom{n}{r} = \binom{n}{n-r}$. Interprete este resultado.

d $\displaystyle\sum_{i=0}^{n} \binom{n}{i} = 2^n$. [*Sugerencia:* considere la expansión binomial de $(x + y)^n$ con $x = y = 1$.]

2.69 Demuestre que $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$.

***2.70** Considere la situación en la que n artículos se han de dividir en $k < n$ subconjuntos distintos. Los coeficientes multinomiales $\binom{n}{n_1\, n_2 \cdots n_k}$ dan el número de particiones distintas donde n_1 artículos están en el grupo 1, n_2 están en el grupo 2, … , n_k están en el grupo k. Demuestre que el número total de particiones distintas es igual a k^n. [*Sugerencia:* recuerde el Ejercicio 2.68(d).]

2.7 Probabilidad condicional y la independencia de eventos

La probabilidad de un evento en ocasiones depende de si sabemos que han ocurrido otros eventos. Por ejemplo, los pescadores deportivos de Florida están vitalmente interesados en la probabilidad de lluvia. La probabilidad de lluvia en un día determinado, si no se hace caso de las condiciones atmosféricas diarias o de otros eventos, es la fracción de días en los que hay lluvia en un largo periodo. Ésta es la *probabilidad incondicional* del evento "lluvia en un día determinado". Ahora suponga que deseamos considerar la probabilidad de lluvia mañana.

Ha llovido casi continuamente durante dos días seguidos y una tormenta tropical se aproxima por la costa. Tenemos información adicional en relación con si llueve o no llueve mañana y estamos interesados en la probabilidad *condicional* de que lloverá *dada* esta información. Un residente de Florida nos diría que la probabilidad condicional de lluvia (dado que ha llovido dos días antes y que se ha pronosticado una tormenta tropical) es mucho mayor que una probabilidad incondicional de lluvia.

La probabilidad incondicional de un 1 en el tiro de un dado balanceado es 1/6. Si sabemos que ha caído un número impar, el número del dado debe ser 1, 3 o 5 y la frecuencia relativa de que haya un 1 es 1/3. La probabilidad condicional de un evento es la probabilidad (frecuencia relativa de ocurrencia) del evento dado el hecho de que uno o más eventos ya han ocurrido. Un examen cuidadoso de este ejemplo indicará el acuerdo de la siguiente definición con el concepto de probabilidad de frecuencia relativa.

DEFINICIÓN **2.9** La *probabilidad condicional de un evento* A, dado que un evento B ha ocurrido, es igual a

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

siempre que $P(B) > 0$. [El símbolo $P(A|B)$ se lee "probabilidad de A dada B".]

Una confirmación adicional de la consistencia de la Definición 2.9 con el concepto de probabilidad de frecuencia relativa se puede obtener de la siguiente construcción. Suponga que un experimento se repite un gran número de veces, N, resulta en A y B, $A \cup B$, n_{11} veces; A y no B, $A \cap \overline{B}$, n_{21} veces; B y no A, $\overline{A} \cap B$, n_{12} veces; y ni A ni B, $\overline{A} \cap \overline{B}$, n_{22}, veces. Estos resultados están contenidos en la Tabla 2.1.

Nótese que $n_{11} + n_{12} + n_{21} + n_{22} = N$. Entonces se deduce que

$$P(A) \approx \frac{n_{11} + n_{21}}{N}, \qquad P(B) \approx \frac{n_{11} + n_{12}}{N}, \qquad P(A|B), \approx \frac{n_{11}}{n_{11} + n_{12}},$$

$$P(B|A) \approx \frac{n_{11}}{n_{11} + n_{21}}, \qquad \text{y } P(A \cap B) \approx \frac{n_{11}}{N},$$

donde $\approx$ se lee *aproximadamente igual a.*

Con estas probabilidades, es fácil ver que

$$P(B|A) \approx \frac{P(A \cap B)}{P(A)} \qquad \text{y} \qquad P(A|B) \approx \frac{P(A \cap B)}{P(B)}.$$

Por tanto, la Definición 2.9 es consistente con el concepto de probabilidad de frecuencia relativa.

Tabla 2.1 Tabla para eventos *A* y *B*

	A	$\overline{A}$	
B	n_{11}	n_{12}	$n_{11} + n_{12}$
$\overline{B}$	n_{21}	n_{22}	$n_{21} + n_{22}$
	$n_{11} + n_{21}$	$n_{12} + n_{22}$	N

EJEMPLO **2.14** Suponga que un dado balanceado se tira una vez. Use la Definición 2.9 para hallar la probabilidad de un 1, dado que se obtuvo un número impar.

Solución Defina estos eventos:

A: observar un 1.

B: observar un número impar.

Buscamos la probabilidad de A dado que el evento B ha ocurrido. El evento $A \cap B$ requiere que se observen un 1 y un número impar. En este caso, $A \subset B$, de modo que $A \cap B = A$ y $P(A \cap B) = P(A) = 1/6$. También, $P(B) = 1/2$ y, usando la Definición 2.9,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Nótese que este resultado está en completo acuerdo con nuestra anterior evaluación intuitiva de esta probabilidad. ■

Suponga que la probabilidad de que ocurra un evento A no es afectada por que ocurra o no ocurra el evento B. Cuando esto sucede, estaríamos inclinados a decir que los eventos A y B son independientes. Esta relación de eventos está expresada por la siguiente definición.

DEFINICIÓN **2.10** Se dice que dos eventos A y B son *independientes* si se cumple cualquiera de los siguientes casos:

$$P(A|B) = P(A),$$
$$P(B|A) = P(B),$$
$$P(A \cap B) = P(A)P(B).$$

De otro modo, se dice que los eventos son *dependientes*.

La noción de independencia como concepto probabilístico está de acuerdo con nuestro uso diario de la palabra si cuidadosamente consideramos los eventos en cuestión. Casi todos estaríamos de acuerdo en que "fumar" y "contraer cáncer de pulmón" no son eventos independientes y de modo intuitivo sentiríamos que la probabilidad de contraer cáncer de pulmón, dado que una persona fuma, es mayor que la probabilidad (incondicional) de contraer cáncer de pulmón. En contraste, los eventos "lloverá hoy" y "lloverá dentro de un mes" muy bien pueden ser independientes.

EJEMPLO **2.15** Considere los siguientes eventos en el tiro de un solo dado:

A: observar un número impar.

B: observar un número par.

C: observar un 1 o 2.

a ¿A y B son eventos independientes?
b ¿A y C son eventos independientes?

Solución **a** Para decidir si A y B son independientes, debemos ver si satisfacen las condiciones de la Definición 2.10. En este ejemplo, $P(A) = 1/2$, $P(B) = 1/2$ y $P(C) = 1/3$. Como $A \cap B = \emptyset$, $P(A|B) = 0$ y es evidente que $P(A|B) \neq P(A)$. Los eventos A y B son dependientes.

b ¿A y C son independientes? Observe que $P(A|C) = 1/2$ y, como antes, $P(A) = 1/2$. Por tanto, $P(A|C) = P(A)$ y A y C son independientes. ■

EJEMPLO **2.16** Tres marcas de café, X, Y y Z, van a ser clasificadas por un juez de acuerdo con su sabor. Defina los siguientes eventos:

A: la marca X es preferida a la Y.
B: la marca X es clasificada como la mejor.
C: la marca X es clasificada como la segunda mejor.
D: la marca X es clasificada como la tercera mejor.

Si el juez en realidad no tiene preferencia por el sabor y al azar asigna lugar a las marcas, ¿el evento A es independiente de los eventos B, C y D?

Solución Los seis puntos muestrales igualmente probables para este experimento están dados por

$$\begin{array}{lll} E_1: XYZ, & E_3: YXZ, & E_5: ZXY, \\ E_2: XZY, & E_4: YZX, & E_6: ZYX, \end{array}$$

donde XYZ denota que X es clasificada como la mejor, Y es la segunda mejor y Z al último.

Entonces $A = \{E_1, E_2, E_5\}$, $B = \{E_1, E_2\}$, $C = \{E_3, E_5\}$, $D = \{E_4, E_6\}$ y se deduce que

$$P(A) = 1/2, \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} = 1, \qquad P(A|C) = 1/2, \qquad P(A|D) = 0.$$

Así, los eventos A y C son independientes, pero los eventos A y B son dependientes. Los eventos A y D también son dependientes. ■

Ejercicios

2.71 Si dos eventos, A y B, son tales que $P(A) = .5$, $P(B) = .3$, y $P(A \cap B) = .1$, encuentre lo siguiente:

a $P(A|B)$
b $P(B|A)$
c $P(A|A \cup B)$

d $P(A|A \cap B)$

e $P(A \cap B|A \cup B)$

2.72 Para cierta población de empleados, el porcentaje que aprueba o reprueba un examen de competencia en un trabajo, indicado de acuerdo con el género, fueron como se ve en la tabla siguiente. Esto es, de todas las personas que tomaron el examen, 24% estaban en la categoría de hombre-aprobado, 16% estuvieron en la categoría de mujer-reprobada y así sucesivamente. Un empleado va a ser seleccionado al azar de esta población. Sea *A* el evento de que el empleado obtenga una calificación de aprobado en el examen y sea *M* el evento de que se seleccione un hombre.

	Género		
Resultado	Masculino (M)	Femenino (F)	*Total*
Aprobado (A)	24	36	60
Reprobado ($\overline{A}$)	16	24	40
Total	40	60	100

a ¿Los eventos *A* y *M* son independientes?

b ¿Los eventos $\overline{A}$ y *F* son independientes?

2.73 Gregor Mendel fue un monje que, en 1865, sugirió una teoría de la herencia basada en la ciencia de la genética. Él identificó individuos heterocigotos para color de flor que tenían dos alelos (un r = alelo color blanco recesivo y un R = alelo color rojo dominante). Cuando estos individuos se apareaban, se observaba que 3/4 de la descendencia tenían flores rojas y 1/4 tenían flores blancas. La tabla siguiente resume este apareamiento; cada padre da uno de sus alelos para formar el gen de la descendencia.

	Padre 2	
Padre 1	r	R
r	rr	rR
R	Rr	RR

Suponemos que es igualmente probable que cada padre contribuye con cualquiera de los dos alelos y que, si uno o dos de los alelos en un par es dominante (R), la descendencia tendrá flores rojas. ¿Cuál es la probabilidad de que un descendiente tenga

a al menos un alelo dominante?,

b al menos un alelo recesivo?,

c un alelo recesivo, dado que la descendencia tiene flores rojas?

2.74 Cien adultos fueron entrevistados en una encuesta por teléfono. De interés fue la opinión de ellos con respecto a las cargas que implican los préstamos para estudiantes universitarios y si quienes respondieron tenían un hijo actualmente en la universidad. Sus respuestas se resumen en la tabla siguiente:

	Carga de préstamo			
Hijo en la universidad	Demasiado alta (A)	Razonable (B)	Demasiado baja (C)	*Total*
Sí (D)	.20	.09	.01	.30
No (E)	.41	.21	.08	.70
Total	.61	.30	.09	1.00

De los siguientes eventos, ¿cuáles son independientes?

a A y D

b B y D

c C y D

2.75 Se reparten cartas, una a la vez, de una baraja de 52 cartas.

a Si las primeras 2 cartas son espadas, ¿cuál es la probabilidad de que las siguientes 3 cartas también sean espadas?

b Si las primeras 3 cartas son todas de espadas, ¿cuál es la probabilidad de que las 2 cartas siguientes sean también espadas?

c Si las primeras 4 cartas son todas de espadas, ¿cuál es la probabilidad de que la siguiente carta sea también una espada?

2.76 Una encuesta de consumidores en una comunidad particular mostró que 10% no estaban satisfechos con los trabajos de plomería realizados en sus casas. La mitad de las quejas se refería al plomero A, que realiza 40% de los trabajos de plomería de la población. Encuentre la probabilidad de que un consumidor obtenga

a un trabajo de plomería no satisfactorio, dado que el plomero era A.

b un trabajo de plomería satisfactorio, dado que el plomero era A.

2.77 Un estudio de conducta, después de un tratamiento hecho a gran número adictos a las drogas, sugiere que la probabilidad de hallarlos culpables no más de dos años después del tratamiento depende de la educación de los infractores. Las proporciones del número total de casos que caen en cuatro categorías de educación-culpabilidad se muestran en la tabla siguiente:

	Situación en no más de 2 años después de tratamiento		
Educación	Culpable	No culpable	*Total*
10 años o más	.10	.30	.40
9 años o menos	.27	.33	.60
Total	.37	.63	1.00

Suponga que se selecciona un solo infractor del programa de tratamiento. Defina los eventos:

A: el infractor tiene 10 o más años de educación.

B: el infractor es hallado culpable en no más de dos años después de terminar el tratamiento.

Encuentre lo siguiente:

a $P(A)$.

b $P(B)$.

c $P(A \cap B)$.

d $P(A \cup B)$.

e $P(\overline{A})$.

f $P(\overline{A \cup B})$.

g $P(\overline{A \cap B})$.

h $P(A|B)$.

i $P(B|A)$.

2.78 En la definición de la independencia de dos eventos, nos dan tres igualdades para comprobarlas: $P(A|B) = P(A)$ o $P(B|A) = P(B)$ o $P(A \cap B) = P(A)P(B)$. Si una cualquiera de estas igualdades se cumple, A y B son independientes. Demuestre que si cualquiera de estas igualdades se cumple, las otras dos también se cumplen.

2.79 Suponga que A y B son eventos mutuamente excluyentes, con $P(A) > 0$ y $P(B) < 1$. ¿A y B son independientes? Justifique su respuesta.

2.80 Suponga que $A \subset B$ y que $P(A) > 0$ y $P(B) > 0$. ¿A y B son independientes? Justifique su respuesta.

2.81 Si $P(A) > 0$, $P(B) > 0$ y $P(A) < P(A|B)$, demuestre que $P(B) < P(B|A)$.

2.82 Suponga que $A \subset B$ y que $P(A) > 0$ y $P(B) > 0$. Demuestre que $P(B|A) = 1$ y $P(A|B) = P(A)/P(B)$.

2.83 Si A y B son eventos mutuamente excluyentes y $P(B) > 0$, demuestre que

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)}.$$

2.8 Dos leyes de probabilidad

Las dos leyes siguientes proporcionan las probabilidades de uniones e intersecciones de eventos. Como tales, desempeñan un importante papel en el método de composición de evento para la solución de problemas de probabilidad.

TEOREMA **2.5**

Ley multiplicativa de probabilidad La probabilidad de la intersección de dos eventos A y B es

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B). \end{aligned}$$

Si A y B son independientes, entonces

$$P(A \cap B) = P(A)P(B).$$

Demostración

La ley multiplicativa se deduce directamente de la Definición 2.9, la definición de probabilidad condicional.

Observe que la ley multiplicativa se puede extender para hallar la probabilidad de la intersección de cualquier número de eventos. Entonces, aplicando dos veces el Teorema 2.5, obtenemos

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] = P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B). \end{aligned}$$

La probabilidad de la intersección de cualquier número de k eventos, por ejemplo, se puede obtener en la misma forma:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &\quad \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}). \end{aligned}$$

La ley aditiva de probabilidad da la probabilidad de la unión de dos eventos.

TEOREMA **2.6**

Ley aditiva de probabilidad La probabilidad de la unión de dos eventos A y B es

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

Si A y B son eventos mutuamente excluyentes, $P(A \cap B) = 0$ y

$$P(A \cup B) = P(A) + P(B).$$

Demostración

La prueba de la ley aditiva se puede deducir al inspeccionar el diagrama de Venn de la Figura 2.10.

Observe que $A \cup B = A \cup (\overline{A} \cap B)$, donde A y $(\overline{A} \cap B)$ son eventos mutuamente excluyentes. Además, $B = (\overline{A} \cap B) \cup (A \cap B)$, donde $(A \cap \overline{B})$ y $(A \cap B)$ son eventos mutuamente excluyentes. Entonces, por el Axioma 3,

$$P(A \cup B) = P(A) + P(\overline{A} \cap B) \quad \text{y} \quad P(B) = P(\overline{A} \cap B) + P(A \cap B).$$

La igualdad dada a la derecha implica que $P(\overline{A} \cap B) = P(B) - P(A \cap B)$. Sustituyendo esta expresión por $P(\overline{A} \cap B)$ en la expresión para $P(A \cup B)$ dada en la ecuación del lado izquierdo del par anterior, obtenemos el resultado deseado:

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

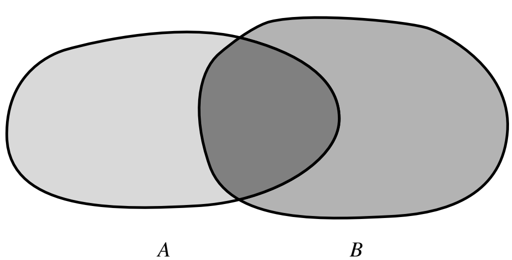
La probabilidad de la unión de tres eventos se puede obtener al hacer uso del Teorema 2.6. Observe que

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

porque $(A \cap B) \cap (A \cap C) = A \cap B \cap C$.

Otro resultado útil que expresa la relación entre la probabilidad de un evento y su complemento se encuentra de inmediato disponible en los axiomas de probabilidad.

FIGURA **2.10**
Diagrama de Venn para la unión de A y B



TEOREMA **2.7** Si A es un evento, entonces

$$P(A) = 1 - P(\overline{A}).$$

Demostración Observe que $S = A \cup \overline{A}$. Como A y $\overline{A}$ son eventos mutuamente excluyentes, se deduce que $P(S) = P(A) + P(\overline{A})$. Por tanto, $P(A) + P(\overline{A}) = 1$ y el resultado sigue.

Como veremos en la Sección 2.9, en ocasiones es más fácil calcular $P(\overline{A})$ que calcular $P(A)$. En tales casos, es más sencillo hallar $P(A)$ por la relación $P(A) = 1 - P(\overline{A})$ que encontrar $P(A)$ directamente.

Ejercicios

2.84 Si A_1, A_2 y A_3 son tres eventos y $P(A_1 \cap A_2) = P(A_1 \cap A_3) \neq$ pero $P(A_2 \cap A_3) = 0$, demuestre que

$$P(\text{al menos una } A_i) = P(A_1) + P(A_2) + P(A_3) - 2P(A_1 \cap A_2).$$

2.85 Si A y B son eventos independientes, demuestre que A y $\overline{B}$ son también independientes. ¿$\overline{A}$ y $\overline{B}$ son independientes?

2.86 Suponga que A y B son dos eventos tales que $P(A) = .8$ y $P(B) = .7$.

a ¿Es posible que $P(A \cap B) = .1$? ¿Por qué sí o por qué no?

b ¿Cuál es el valor más pequeño posible de $P(A \cap B)$?

c ¿Es posible que $P(A \cap B) = .77$? ¿Por qué sí o por qué no?

d ¿Cuál es el máximo valor posible para $P(A \cap B)$?

2.87 Suponga que A y B son dos eventos tales que $P(A) + P(B) > 1$.

a ¿Cuál es el mínimo valor posible para $P(A \cap B)$?

b ¿Cuál es el máximo valor posible para $P(A \cap B)$?

2.88 Suponga que A y B son dos eventos tales que $P(A) = .6$ y $P(B) = .3$.

a ¿Es posible que $P(A \cap B) = .1$? ¿Por qué sí o por qué no?

b ¿Cuál es el mínimo valor posible para $P(A \cap B)$?

c ¿Es posible que $P(A \cap B) = .7$? ¿Por qué sí o por qué no?

d ¿Cuál es el máximo valor posible para $P(A \cap B)$?

2.89 Suponga que A y B son dos eventos tales que $P(A) + P(B) < 1$.

a ¿Cuál es el mínimo valor posible para $P(A \cap B)$?

b ¿Cuál es el máximo valor posible para $P(A \cap B)$?

2.90 Suponga que hay 1 en 50 probabilidades de lesión en un solo intento de paracaidismo.

a Si suponemos que los resultados de diferentes saltos son independientes, ¿cuál es la probabilidad de que una paracaidista se lesione si salta dos veces?

b Un amigo dice que si hay 1 en 50 probabilidades de lesión en un solo salto entonces hay un 100% de probabilidad de lesión si una paracaidista salta 50 veces. ¿Tiene razón su amigo? ¿Por qué?

2.91 ¿Pueden A y B ser mutuamente excluyentes si $P(A) = .4$ y $P(B) = .7$? ¿Si $P(A) = .4$ y $P(B) = .3$? ¿Por qué?

2.92 Una política que requiere que todos los empleados hospitalarios hagan exámenes de detector de mentiras reduce pérdidas debidas a robos, pero algunos empleados consideran tales exámenes como una violación a sus derechos. Experiencias pasadas indican que los detectores de mentiras tienen porcentajes de precisión que varían de 92% a 99%.[2] Para tener alguna idea de los riesgos a los que se enfrentan los empleados cuando hacen un examen de detector de mentiras, suponga que la probabilidad es .05 de que un detector de mentiras concluya que una persona está mintiendo y que, en realidad, esté diciendo la verdad, y suponga que cualquier par de exámenes son independientes. ¿Cuál es la probabilidad de que una máquina concluya que

a cada uno de tres empleados está mintiendo cuando todos están diciendo la verdad?,

b al menos uno de los tres empleados está mintiendo cuando todos están diciendo la verdad?

2.93 En un juego de *softball*, una jugadora tiene tres oportunidades para conectar un imparable. En cada intento, ella conecta un *hit*, H, o no conecta, M. El juego requiere que la jugadora debe alternar la mano que usa en intentos sucesivos, es decir, si ella hace su primer intento con la mano derecha, debe usar la izquierda para el segundo intento y su mano derecha para el tercero. Su oportunidad de conectar un *hit* con la derecha es .7 y con la izquierda es .4. Suponga que los resultados de intentos sucesivos son independientes y que ella gana el juego si conecta al menos dos *hits* consecutivos. Si ella hace su primer intento con la mano derecha, ¿cuál es la probabilidad de que gane el juego?

2.94 Un sistema detector de humo utiliza dos dispositivos, A y B. Si hay humo, la probabilidad que sea detectado por el dispositivo A es .95; por el dispositivo B, .90; y por ambos dispositivos, .88.

a Si hay humo, encuentre la probabilidad de que el humo sea detectado ya sea por el dispositivo A o el B o por ambos.

b Encuentre la probabilidad de que el humo no sea detectado.

2.95 Dos eventos A y B son tales que $P(A) = .2$, $P(B) = .3$ y $P(A \cup B) = .4$. Encuentre lo siguiente:

a $P(A \cap B)$
b $P(\overline{A} \cup \overline{B})$
c $P(\overline{A} \cap \overline{B})$
d $P(\overline{A}|B)$

2.96 Si A y B son eventos independientes con $P(A) = .5$ y $P(B) = .2$, encuentre lo siguiente:

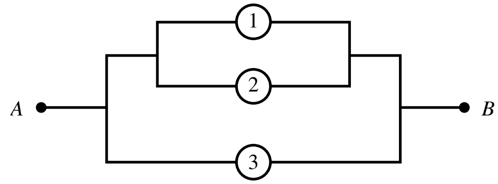
a $P(A \cup B)$
b $P(\overline{A} \cap \overline{B})$
c $P(\overline{A} \cup \overline{B})$

2.97 Considere la porción siguiente de un circuito eléctrico con tres relevadores. Circulará corriente del punto a al b si hay al menos un circuito cerrado cuando los relevadores están activados. Los relevadores funcionan mal y no cierran cuando se activan. Suponga que los relevadores actúan de manera independiente entre sí y cierran en forma apropiada cuando son activados con una probabilidad de .9.

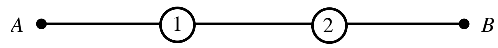
a ¿Cuál es la probabilidad de que circule corriente cuando los relevadores se activan?

b Dado que circuló corriente cuando se activaron los relevadores, ¿cuál es la probabilidad de que el relevador 1 funcione?

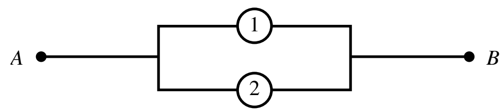
2. *Fuente:* Copyright © 1980 Sentinel Communications Co. Todos los derechos reservados.



2.98 Con relevadores operando como en el Ejercicio 2.97 compare la probabilidad de que circule corriente de a a b en el sistema en serie que se muestra



con la probabilidad de circulación en el sistema en paralelo que se muestra.



2.99 Suponga que A y B son eventos independientes tales que la probabilidad de que ninguno suceda es a y la probabilidad de B es b. Demuestre que $P(A) = \dfrac{1-b-a}{1-b}$.

***2.100** Demuestre que el Teorema 2.6, la ley aditiva de probabilidad, se cumple para probabilidades *condicionales.* Esto es, si A, B y C son eventos tales que $P(C) > 0$, demuestre que $P(A \cup B|C) = P(A|C) +$ $.P(B|C) - P(A \cap B|C)$. [*Sugerencia:* haga uso de la ley distributiva $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.]

2.101 Los artículos que pasan por una línea de inspección son revisados visualmente por dos inspectores sucesivos. Cuando un artículo defectuoso pasa por la línea de inspección, la probabilidad de que sea captado por el primer inspector es .1. El segundo inspector "no ve" cinco de entre diez de los artículos defectuosos que deja pasar el primer inspector. ¿Cuál es la probabilidad de que un artículo defectuoso no sea captado por ambos inspectores?

2.102 Las enfermedades I y II son prevalecientes entre las personas de cierta población. Se supone que 10% de la población contraerá la enfermedad I alguna vez en su vida, 15% contraerá la enfermedad II con el tiempo y 3% contraerá ambas enfermedades.

a Encuentre la probabilidad de que una persona escogida al azar de esta población contraiga al menos una enfermedad.

b Encuentre la probabilidad condicional de que una persona escogida al azar de esta población contraiga ambas enfermedades, dado que él o ella ha contraído al menos una enfermedad.

2.103 Consulte el Ejercicio 2.50. Horas después de que fuera anunciada la manipulación de la lotería del estado de Pennsylvania, los oficiales de la lotería del estado de Connecticut quedaron sorprendidos al enterarse que su número ganador para el día fue el 666 (*Los Angeles Times*, 21 de septiembre de 1980).

a Toda la evidencia indica que la selección del 666 de Connecticut fue pura casualidad. ¿Cuál es la probabilidad de que un 666 salga en Connecticut, dado que un 666 había sido seleccionado el 24 de abril de 1980 en la lotería de Pennsylvania?

b ¿Cuál es la probabilidad de sacar un 666 en la lotería de Pennsylvania el 24 de abril de 1980 (recuerde, este resultado fue una manipulación) y un 666 el 19 de septiembre de 1980 en la lotería de Connecticut?

2.104 Si A y B son dos eventos, demuestre que $P(A \cap B) \geq 1 - P(\overline{A}) - P(\overline{B})$. [*Nota:* ésta es una versión simplificada de la *desigualdad de Bonferroni*.]

2.105 Si la probabilidad de lesiones en cada salto en paracaídas de un individuo es .05, use el resultado del Ejercicio 2.104 para dar un límite inferior para la probabilidad de un aterrizaje seguro en los dos saltos.

2.106 Si A y B son eventos igualmente probables y requerimos que la probabilidad de su intersección sea al menos .98, ¿cuál es $P(A)$?

2.107 Sean A, B y C eventos tales que $P(A) > P(B)$ y $P(C) > 0$. Construya un ejemplo para demostrar que es posible que $P(A|C) < P(B|C)$.

2.108 Si A, B y C son tres eventos, use dos aplicaciones del resultado del Ejercicio 2.104 para demostrar que $P(A \cap B \cap C) \geq 1 - P(\overline{A}) - P(\overline{B}) - P(\overline{C})$.

2.109 Si A, B y C son tres eventos igualmente probables, ¿cuál es el mínimo valor para $P(A)$ tal que $P(A \cap B \cap C)$ siempre sea mayor que 0.95?

2.9 Cálculo de la probabilidad de un evento: el método de composición de evento

Aprendimos en la Sección 2.4 que los conjuntos (eventos) se pueden expresar en ocasiones como uniones, intersecciones o complementos de otros conjuntos. El método de composición de evento para calcular la probabilidad de un evento, A, lo expresa como una composición que comprende uniones y/o intersecciones de otros eventos. Las leyes de probabilidad se aplican entonces para hallar $P(A)$. Ilustraremos este método con un ejemplo.

EJEMPLO **2.17** De los votantes en una ciudad, 40% son republicanos y 60% son demócratas. Entre los republicanos, 70% están a favor de una emisión de bonos, en tanto que 80% de los demócratas están a favor de la emisión. Si un votante se selecciona al azar en la ciudad, ¿cuál es la probabilidad de esté a favor de la emisión de bonos?

Solución Denote con F el evento "a favor de la emisión de bonos", con R el evento "se selecciona un republicano" y con D el evento "se selecciona un demócrata". Entonces $P(R) = .4$, $P(D) = .6$, $P(F|R) = .7$ y $P(F|D) = .8$. Nótese que

$$P(F) = P[(F \cap R) \cup (F \cap D)] = P(F \cap R) + P(F \cap D)$$

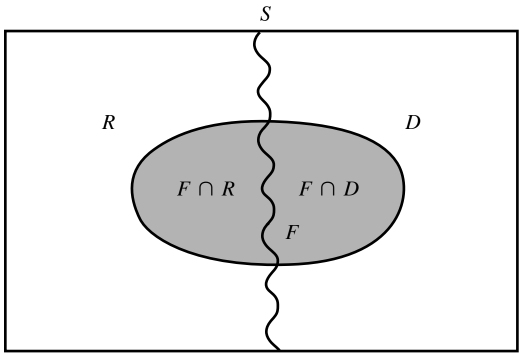
porque $(F \cap R)$ y $(F \cap D)$ son eventos mutuamente excluyentes. La Figura 2.11 ayudará a visualizar el resultado de que $F = (F \cap R) \cup (F \cap D)$. Ahora

$$P(F \cap R) = P(F|R)P(R) = (.7)(.4) = .28,$$
$$P(F \cap D) = P(F|D)P(D) = (.8)(.6) = .48.$$

Se deduce que

$$P(F) = .28 + .48 = .76.$$

FIGURA **2.11** Diagrama de Venn para eventos del Ejemplo 2.17



■

EJEMPLO **2.18** En el Ejemplo 2.7 consideramos un experimento en el que se registraron los cumpleaños de 20 personas seleccionadas al azar. En ciertas condiciones encontramos que $P(A) = .5886$, donde A denota el evento de que cada persona tenga un cumpleaños diferente. Denote con B el evento de que al menos un par de personas comparta cumpleaños. Encuentre $P(B)$.

Solución El evento B es el conjunto de todos los puntos muestrales en S que no están en A, es decir, $B = \overline{A}$. Por tanto,

$$P(B) = 1 - P(A) = 1 - .5886 = .4114.$$

(Casi todos estaríamos de acuerdo en que esta probabilidad es sorprendentemente alta.) ■

Consultemos el Ejemplo 2.4, que involucra los dos jugadores de tenis y denotemos con D_1 y D_2 los eventos de que el jugador A gane el primero y segundo juegos, respectivamente. La información dada en el ejemplo implica que $P(D_1) = P(D_2) = 2/3$. Además, si hacemos la suposición de que D_1 y D_2 son independientes, se deduce que $P(D_1 \cap D_2) = 2/3 \times 2/3 = 4/9$. En ese ejemplo identificamos el evento simple E_1, que denotamos como AA queriendo decir que el jugador A ganó ambos juegos. Con la notación presente,

$$E_1 = D_1 \cap D_2,$$

y por tanto $P(E_1) = 4/9$. Las probabilidades asignadas a los otros eventos simples del Ejemplo 2.4 se pueden verificar de un modo similar.

El método de composición de evento no tendrá éxito a menos que se conozcan las probabilidades de los eventos que aparecen en $P(A)$ (después de aplicar las leyes aditiva y multiplicativa). Si no se conoce una o más de estas probabilidades, el método falla. En ocasiones es deseable formar composiciones de eventos mutuamente excluyentes o independientes; éstos simplifican el uso de la ley aditiva, y la ley multiplicativa de probabilidad es más fácil de aplicar a eventos independientes.

A continuación veamos un resumen de los pasos seguidos en el método de composición de evento:

1. Definir el experimento.
2. Visualizar la naturaleza de los puntos muestrales. Identificar unos pocos para aclarar el modo de pensar del experto en estadística.
3. Escribir una ecuación que exprese el evento de interés, A por ejemplo, como una composición de dos o más eventos, usando uniones, intersecciones y/o complementos. (Nótese que esto iguala conjuntos de puntos.) Asegurarse que el evento A y el evento implicado por la composición representan el mismo conjunto de puntos muestrales.
4. Aplicar las leyes aditiva y multiplicativa de probabilidad a las composiciones obtenidas en el paso 3 para hallar $P(A)$.

El paso 3 es el más difícil porque podemos formar muchas composiciones que serán equivalentes al evento A. El truco es formar una composición en la que se conozcan todas las probabilidades que aparezcan en el paso 4.

El método de composición de evento no requiere indicar los puntos muestrales en S, pero requiere un claro entendimiento de la naturaleza de un punto muestral típico. El principal error que los estudiantes cometen al aplicar el método de composición de evento es cuando escriben la composición. Esto es, la ecuación del conjunto de puntos que expresa A como la unión y/o intersección de otros eventos es, con frecuencia, incorrecta. Siempre compruebe su igualdad y asegúrese de que la composición implica un evento que contenga el mismo conjunto de puntos muestrales que en A.

Una comparación de los métodos de punto muestral y de composición de evento para calcular la probabilidad de un evento se puede obtener al aplicar ambos métodos al mismo problema. Aplicaremos el método de composición de evento al problema de seleccionar solicitantes que fue resuelto por el método de punto muestral en los Ejemplos 2.11 y 2.12.

EJEMPLO **2.19** Dos solicitantes se seleccionan al azar de entre cinco que han solicitado un trabajo. Encuentre la probabilidad de que se seleccione exactamente uno de los dos mejores, evento A.

Solución Defina los siguientes dos eventos:

B: Sacar al mejor y a uno de los tres solicitantes más malos.
C: Sacar al segundo mejor y a uno de los tres solicitantes más malos.

Los eventos B y C son mutuamente excluyentes y $A = B \cup C$. También, sea $D_1 = B_1 \cap B_2$, donde

B_1 = Sacar al mejor en la primera oportunidad,
B_2 = Sacar uno de los tres solicitantes más malos en el segundo intento,

y $D_2 = B_3 \cap B_4$, donde

B_3 = Sacar uno de los tres solicitantes más malos en el primer intento,
B_4 = Sacar al mejor en el segundo intento.

Observe que $B = D_1 \cup D_2$.

Del mismo modo, sea $G_1 = C_1 \cap C_2$ y $G_2 = C_3 \cap C_4$, donde C_1, C_2, C_3 y C_4 están definidas como B_1, B_2, B_3 y B_4, con las palabras *segundo mejor* sustituyendo a *mejor*. Nótese que D_1 y D_2 y G_1 y G_2 son pares de eventos mutuamente excluyentes y que

$$A = B \cup C = (D_1 \cup D_2) \cup (G_1 \cup G_2),$$
$$A = (B_1 \cap B_2) \cup (B_3 \cap B_4) \cup (C_1 \cap C_2) \cup (C_3 \cap C_4).$$

Si aplicamos la ley aditiva de la probabilidad a estos cuatro eventos mutuamente excluyentes tenemos

$$P(A) = P(B_1 \cap B_2) + P(B_3 \cap B_4) + P(C_1 \cap C_2) + P(C_3 \cap C_4).$$

Al aplicar la ley multiplicativa tenemos

$$P(B_1 \cap B_2) = P(B_1)P(B_2|B_1).$$

La probabilidad de sacar al mejor en el primer intento es

$$P(B_1) = 1/5.$$

Del mismo modo, la probabilidad de sacar uno de los tres más malos en el segundo intento, dado que el mejor fue sacado en la primera selección, es

$$P(B_2|B_1) = 3/4.$$

Entonces

$$P(B_1 \cap B_2) = P(B_1)P(B_2|B_1) = (1/5)(3/4) = 3/20.$$

Las probabilidades de todas las otras intersecciones en $P(A)$, $P(B_3 \cap B_4)$, $P(C_1 \cap C_2)$ y $P(C_3 \cap C_4)$ se obtienen exactamente del mismo modo y todas son iguales a 3/20. Entonces

$$\begin{aligned} P(A) &= P(B_1 \cap B_2) + P(B_3 \cap B_4) + P(C_1 \cap C_2) + P(C_3 \cap C_4) \\ &= (3/20) + (3/20) + (3/20) + (3/20) = 3/5. \end{aligned}$$

Esta respuesta es idéntica a la obtenida en el Ejemplo 2.12, donde $P(A)$ se calculó usando el método de punto muestral. ■

EJEMPLO **2.20** Se sabe que un paciente con una enfermedad responderá al tratamiento con probabilidad igual a .9. Si tres pacientes con la enfermedad son tratados y responden de manera independiente, encuentre la probabilidad de que al menos uno responda.

Solución Defina los siguientes eventos:

A: Al menos uno de los tres pacientes responderá,
B_1: El primer paciente no responderá.
B_2: El segundo paciente no responderá.
B_3: El tercer paciente no responderá.

Entonces observe que $\overline{A} = B_1 \cap B_2 \cap B_3$. El teorema 2.7 implica que

$$\begin{aligned} P(A) &= 1 - P(\overline{A}) \\ &= 1 - P(B_1 \cap B_2 \cap B_3). \end{aligned}$$

Aplicando la ley multiplicativa tenemos

$$P(B_1 \cap B_2 \cap B_3) = P(B_1)P(B_2|B_1)P(B_3|B_1 \cap B_2),$$

donde, debido a que los eventos son independientes,

$$P(B_2|B_1) = P(B_2) = 0.1 \quad \text{y} \quad P(B_3|B_1 \cap B_2) = P(B_3) = 0.1.$$

Sustituyendo $P(B_i) = .1$, $i = 1, 2, 3$, obtenemos

$$P(A) = 1 - (.1)^3 = .999.$$

Observe que hemos demostrado la utilidad de eventos complementarios. Este resultado es importante porque con frecuencia es más fácil hallar la probabilidad del complemento, $P(\overline{A})$, que hallar $P(A)$ directamente. ■

EJEMPLO **2.21** La observación de una fila de espera en una clínica médica indica que la probabilidad de que una nueva llegada sea un caso de emergencia es $p = 1/6$. Encuentre la probabilidad de que el r-ésimo paciente sea el primer caso de emergencia. (Suponga que las condiciones en que llegan los pacientes representan eventos independientes.)

Solución El experimento consiste en observar las llegadas de pacientes hasta que aparezca el primer caso de emergencia. Entonces los puntos muestrales para el experimento son

E_i: El i-ésimo paciente es el primer caso de emergencia, para $i = 1, 2, \ldots$.

Como sólo un punto muestral cae en el evento de interés,

$$P(r\text{-ésimo paciente es el primer caso de emergencia}) = P(E_r).$$

Ahora defina A_i para denotar el evento de que la i-ésima llegada no sea un caso de emergencia. Entonces podemos representar E_r como la intersección

$$E_r = A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{r-1} \cap \overline{A}_r.$$

Aplicando la ley multiplicativa tenemos

$$P(E_r) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(\overline{A}_r|A_1 \cap \cdots \cap A_{r-1}),$$

y como los eventos $A_1, A_2, \ldots, A_{r-1}$ y $\overline{A}_r$ son independientes, se deduce que

$$\begin{aligned} P(E_r) &= P(A_1)P(A_2) \cdots P(A_{r-1})P(\overline{A}_r) = (1 - p)^{r-1} p \\ &= (5/6)^{r-1}(1/6), \qquad r = 1, 2, 3, \ldots. \end{aligned}$$

Observe que

$$\begin{aligned} P(S) &= P(E_1) + P(E_2) + P(E_3) + \cdots + P(E_i) + \cdots \\ &= (1/6) + (5/6)(1/6) + (5/6)^2(1/6) + \cdots + (5/6)^{i-1}(1/6) + \cdots \\ &= \frac{1}{6}\sum_{i=0}^{\infty}\left(\frac{5}{6}\right)^i = \frac{1/6}{1-(5/6)} = 1. \end{aligned}$$

Este resultado se obtiene de la fórmula para la suma de una *serie geométrica* dada en el Apéndice A1.11. Esta fórmula, que expresa que si $|r| < 1$, $\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$, es útil en muchos problemas simples de probabilidad. ■

EJEMPLO **2.22** Una mona ha de demostrar que reconoce los colores al lanzar al aire una bola roja, una negra y una blanca en cajas de los mismos colores respectivamente, una bola por caja. Si la mona no ha aprendido los colores y simplemente lanza una bola en una caja al azar, encuentre las probabilidades de los siguientes resultados:

a No hay coincidencia de colores.
b Hay exactamente una coincidencia de colores.

Solución Este problema se puede resolver si se hace una lista de puntos muestrales porque sólo hay tres pelotas, pero se ilustrará un método más general. Defina los eventos siguientes:

A_1: Hay una coincidencia de colores en la caja roja.
A_2: Hay una coincidencia de colores en la caja negra.
A_3: Hay una coincidencia de colores en la caja blanca.

Hay $3! = 6$ formas igualmente probables de lanzar pelotas al azar en las cajas con una pelota en cada caja. También, hay sólo $2! = 2$ formas de lanzar las pelotas en las cajas si se requiere que una caja en particular contenga una coincidencia de color. Por tanto,

$$P(A_1) = P(A_2) = P(A_3) = 2/6 = 1/3.$$

Del mismo modo, se deduce que

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = P(A_1 \cap A_2 \cap A_3) = 1/6.$$

Podemos ahora contestar los incisos a y b usando el método de composición de eventos.

a Observe que

$$\begin{aligned} P\,(\text{no hay coincidencia de colores}) &= 1 - P\,(\text{al menos una coincidencia de colores}) \\ &= 1 - P(A_1 \cup A_2 \cup A_3) \\ &= 1 - [P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) \\ &\quad - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)] \\ &= 1 - [3(1/3) - 3(1/6) + (1/6)] = 2/6 = 1/3. \end{aligned}$$

b El estudiante debe demostrar que

$$
\begin{aligned}
P\text{ (exactamente una pareja)} &= P(A_1) + P(A_2) + P(A_3) \\
&\quad - 2[P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)] \\
&\quad + 3[P(A_1 \cap A_2 \cap A_3)] \\
&= (3)(1/3) - (2)(3)(1/6) + (3)(1/6) = 1/2.
\end{aligned}
$$

■

La mejor forma de aprender a resolver problemas de probabilidad es haciéndolos. Para ayudar al estudiante a desarrollar experiencia, al final de esta sección, de este capítulo y de la bibliografía se presentan muchos ejercicios.

Ejercicios

2.110 De los artículos producidos diariamente por una fábrica, 40% provienen de la línea I y 60% de la línea II. La línea I tiene un porcentaje de 8% de piezas defectuosas en tanto que la II tiene un porcentaje de 10%. Si se escoge al azar una pieza de la producción diaria, encuentre la probabilidad de que no esté defectuosa.

2.111 Una agencia de publicidad observa que aproximadamente 1 de cada 50 compradores potenciales de un producto ve el anuncio en una revista determinada y 1 de cada 5 ve un anuncio correspondiente en televisión. Uno de cada 100 los ve a ambos. Uno de cada 3 en realidad compra el producto después de ver el anuncio y 1 de cada 10 sin verlo. ¿Cuál es la probabilidad de que un cliente potencial seleccionado al azar compre el producto?

2.112 Tres aparatos de radar, que operan de modo independiente, se ajustan para detectar cualquier avión que vuele por cierta zona. Cada aparato tiene una probabilidad de .02 de no detectar un avión en su zona. Si un avión entra en ésta, ¿cuál es la probabilidad de que

a no sea detectado?,

b sea detectado por los tres aparatos de radar?

2.113 Considere uno de los radares del Ejercicio 2.112. ¿Cuál es la probabilidad de que detecte correctamente tres aviones antes de no detectar uno de ellos, si las llegadas de aviones son eventos individuales independientes que suceden en tiempos diferentes?

2.114 Un detector de mentiras mostrará una lectura positiva (indica una mentira) 10% del tiempo cuando una persona está diciendo la verdad y 95% del tiempo cuando está mintiendo. Suponga que dos personas son sospechosas en un delito cometido por una persona y (de seguro) una es culpable y mentirá. Suponga, además, que el detector de mentiras opera de manera independiente para la persona honesta y para la mentirosa. ¿Cuál es la probabilidad de que el detector

a muestre una lectura positiva para ambos sospechosos?,

b muestre una lectura positiva para el sospechoso culpable y una lectura negativa para el sospechoso inocente?,

c esté completamente equivocado, es decir, dé una lectura positiva para el sospechoso inocente y una negativa para el culpable?,

d dé una lectura positiva para cualquiera de los sospechosos o para ambos?

2.115 Un equipo de futbol tiene una probabilidad de .75 de ganar cuando juegue con cualquiera de los otros equipos en su conferencia. Si los juegos son independientes, ¿cuál es la probabilidad de que el equipo gane todos los juegos de su conferencia?

2.116 Una red de comunicaciones tiene un sistema integrado de seguridad contra fallas. En este sistema, si falla la línea I, la señal se desvía a la línea II; si también falla la línea II, la señal se desvía a la línea III. La probabilidad de falla de cualquiera de estas tres líneas es .01 y las fallas de estas líneas son eventos independientes. ¿Cuál es la probabilidad de que este sistema de tres líneas no falle por completo?

2.117 Una estación de inspección de autos de un estado tiene dos equipos de inspección. El equipo 1 es poco severo y pasa todos los automóviles de fabricación reciente; el equipo 2 rechaza todos los autos en una primera inspección porque "los faros no están ajustados correctamente". Cuatro automovilistas no enterados llevan sus autos a la estación para ser inspeccionados en cuatro días diferentes y al azar seleccionan uno de los dos equipos.

a Si los cuatro autos son nuevos y en excelentes condiciones, ¿cuál es la probabilidad de que tres de ellos sean rechazados?

b ¿Cuál es la probabilidad de que los cuatro pasen?

2.118 La víctima de un accidente morirá a menos que en los siguientes 10 minutos reciba sangre Rh positiva de tipo A, que puede ser donada por una sola persona. El hospital requiere 2 minutos para clasificar el tipo de sangre de un donador en perspectiva y 2 minutos para completar la transfusión de sangre. Hay numerosos donadores de sangre no clasificada y 40% de ellos tienen sangre Rh positiva tipo A. ¿Cuál es la probabilidad de que la víctima del accidente sea salvada si sólo hay un equipo para clasificar el tipo de sangre? Suponga que el equipo de clasificación es reutilizable pero puede procesar sólo un donador a la vez.

***2.119** Suponga que dos dados balanceados se lanzan repetidamente y la suma de las dos caras superiores se determina en cada tiro. ¿Cuál es la probabilidad de obtener

a una suma de 3 antes de obtener una suma de 7?,

b una suma de 4 antes de obtener una suma de 7?

2.120 Suponga que dos refrigeradores defectuosos se han incluido en un embarque de seis refrigeradores. El comprador empieza a probar los seis refrigeradores, uno por uno.

a ¿Cuál es la probabilidad de que el último refrigerador defectuoso sea encontrado en la cuarta prueba?

b ¿Cuál es la probabilidad de que no más de cuatro refrigeradores necesiten ser probados para localizar ambos refrigeradores defectuosos?

c Cuando nos indican que exactamente uno de los dos refrigeradores ha sido localizado en las primeras dos pruebas, ¿cuál es la probabilidad de que el refrigerador defectuoso restante se encuentre en la tercera o cuarta pruebas?

2.121 Un nuevo secretario ha recibido n contraseñas de computadora, sólo una de las cuales permitirá el acceso a un archivo de computadora. Como el secretario no tiene idea de cuál contraseña es la correcta, selecciona una de ellas al azar y la prueba. Si es incorrecta, la desecha y selecciona al azar otra contraseña de entre las restantes, prosiguiendo así hasta que encuentra la correcta.

a ¿Cuál es la probabilidad de que el secretario obtenga la contraseña correcta en el primer intento?

b ¿Cuál es la probabilidad de que el secretario obtenga la contraseña correcta en el segundo intento? ¿Y en el tercero?

c Un sistema de seguridad se ha iniciado para que si se prueban tres contraseñas incorrectas el archivo de computadora queda bloqueado y se niega el acceso. Si $n = 7$, ¿cuál es la probabilidad de que el secretario tenga acceso al archivo?

2.10 Ley de probabilidad total y regla de Bayes

El método de composición de evento para resolver problemas de probabilidad en ocasiones se facilita al ver el espacio muestral, S, como una unión de subconjuntos mutuamente excluyentes y usar la siguiente *ley de probabilidad total*. Los resultados de esta sección están basados en la siguiente construcción.

DEFINICIÓN **2.11**

Para algún entero positivo k, sean los conjuntos $B_1, B_2, \ldots, B_k$ tales que

1. $S = B_1 \cup B_2 \cup \cdots \cup B_k$.
2. $B_i \cap B_j = \emptyset$, para $i \neq j$.

Entonces se dice que la colección de conjuntos $\{B_1, B_2, \ldots, B_k\}$ es una *partición* de S.

Si A es cualquier subconjunto de S y $\{B_1, B_2, \ldots, B_k\}$ es una partición de S, A puede *descomponerse* como sigue:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k).$$

La Figura 2.12 ilustra esta descomposición para $k = 3$.

TEOREMA **2.8**

Suponga que $\{B_1, B_2, \ldots, B_k\}$ es una partición de S (vea la Definición 2.11) tal que $P(B_i) > 0$, para $i = 1, 2, \ldots, k$. Entonces, para cualquier evento A

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

Demostración

Cualquier subconjunto A de S puede escribirse como

$$\begin{aligned} A &= A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) \\ &= (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k). \end{aligned}$$

Observe que, como $\{B_1, B_2, \ldots, B_k\}$ es una partición de S, si $i \neq j$,

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$$

y que $(A \cap B_i)$ y $(A \cap B_j)$ son eventos mutuamente excluyentes. Entonces,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i). \end{aligned}$$

En los ejemplos y ejercicios que siguen, el lector verá que en ocasiones es mucho más fácil calcular las probabilidades condicionales $P(A|B_i)$ para una B_i escogida de manera apropiada que calcular $P(A)$ directamente. En tales casos, la ley de probabilidad total puede aplicarse

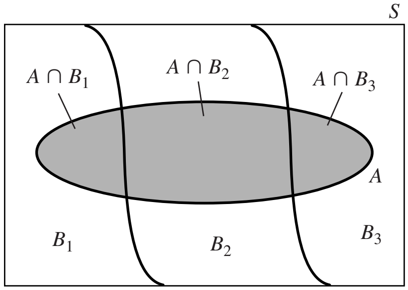


FIGURA **2.12** Descomposición del evento *A*

para determinar $P(A)$. Usando el resultado del Teorema 2.8, es una cuestión sencilla deducir el resultado conocido como *regla de Bayes.*

TEOREMA **2.9** **Regla de Bayes** Suponga que $\{B_1, B_2, \ldots, B_k\}$ es una partición de S (vea la Definición 2.11) tal que $P(B_i) > 0$, para $i = 1, 2, \ldots, k$. Entonces

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Demostración La demostración se deduce directamente de la definición de probabilidad condicional y la ley de probabilidad total. Observe que

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

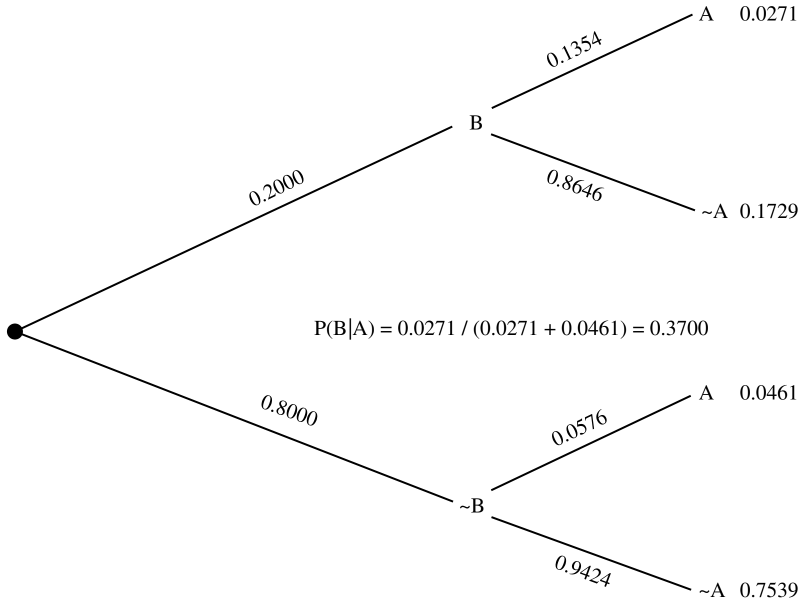
EJEMPLO **2.23** Un fusible electrónico es producido por cinco líneas de producción en una operación de manufactura. Los fusibles son costosos, sumamente confiables y se envían a proveedores en lotes de 100 unidades. Como la prueba es destructiva, la mayoría de los compradores de fusibles prueban sólo un número pequeño de ellos antes de decidirse a aceptar o rechazar lotes de fusibles que lleguen.

Las cinco líneas de producción producen fusibles al mismo ritmo y normalmente producen sólo 2% de fusibles defectuosos, que se dispersan al azar en la producción. Desafortunadamente, la línea 1 de producción sufrió problemas mecánicos y produjo 5% de piezas defectuosas durante el mes de marzo. Esta situación llegó al conocimiento del fabricante después de que los fusibles ya habían sido enviados. Un cliente recibió un lote producido en marzo y probó tres fusibles. Uno falló. ¿Cuál es la probabilidad de que el lote se haya producido en la línea 1? ¿Cuál es la probabilidad de que el lote haya provenido de una de las otras cuatro líneas?

Solución Denotemos con B el evento de que un fusible se sacó de la línea 1 y denotemos con A el evento de que el fusible estuviera defectuoso. Se deduce entonces directamente que

$$P(B) = 0.2 \quad \text{y} \quad (A|B) = 3(.05)(.95)^2 = .135375.$$

FIGURA **2.13** Diagrama de árbol para cálculos en el Ejemplo 2.23. ~A y ~B son notaciones alternativas para $\overline{A}$ y $\overline{B}$, respectivamente.



Del mismo modo,

$$P(\overline{B}) = 0.8 \quad \text{y} \quad P(A|\overline{B}) = 3(.02)(.98)^2 = .057624.$$

Observe que estas probabilidades condicionales fueron muy fáciles de calcular. Usando la ley de probabilidad total,

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \\ &= (.135375)(.2) + (.057624)(.8) = .0731742. \end{aligned}$$

Por último,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{(.135375)(.2)}{.0731742} = .37,$$

y

$$P(\overline{B}|A) = 1 - P(B|A) = 1 - .37 = .63.$$

La Figura 2.13, obtenida usando la aplicación *Bayes' Rule as a Tree*, ilustra los diversos pasos en el cálculo de $P(B|A)$. ■

Ejercicios

2.122 **Ejercicio de aplicación** Use la aplicación *Bayes' Rule as a Tree* para obtener los resultados dados en la Figura 2.13.

2.123 **Ejercicio de aplicación** Consulte el Ejercicio 2.122 y el Ejemplo 2.23. Suponga que las líneas 2 a la 5 siguen iguales, pero la línea 1 fue parcialmente reparada y produjo un menor porcentaje de defectos.

a ¿Qué impacto tendría esto en $P(A|B)$?

b Suponga que $P(A\,|\,B)$ disminuyó a .12 y todas las demás probabilidades siguieron sin cambio. Use la aplicación *Bayes' Rule as a Tree* para reevaluar $P(A\,|\,B)$.

c ¿Cómo se compara la respuesta que usted obtuvo en el inciso b con la obtenida en el Ejercicio 2.122? ¿Le sorprende este resultado?

d Suponga que todas las probabilidades siguen iguales excepto $P(A\,|\,B)$. Use la aplicación de prueba y error para hallar el valor de $P(A\,|\,B)$ para el cual $P(B\,|A)= .3000$.

e Si la línea 1 produjo sólo artículos defectuosos pero todas las otras probabilidades siguieron sin cambio, ¿cuál es $P(B\,|A)$?

f Un amigo esperaba que la respuesta al inciso e fuera 1. Explique por qué, dadas las condiciones del inciso e, $P(B\,|A) \neq 1$.

2.124 Una población de electores contiene 40% de republicanos y 60% de demócratas. Se publica que 30% de los republicanos y 70% de los demócratas están a favor de un tema de elección. Se encuentra que una persona seleccionada al azar de esta población está a favor del tema en cuestión. Encuentre la probabilidad condicional de que esta persona sea un demócrata.

2.125 Una prueba de diagnóstico para una enfermedad es tal que (correctamente) detecta la enfermedad en 90% de los individuos que en realidad tienen la enfermedad. También, si una persona no tiene la enfermedad, la prueba reportará que él o ella no la tiene con probabilidad .9. Sólo 1% de la población tiene la enfermedad en cuestión. Si una persona es seleccionada al azar de la población y la prueba de diagnóstico indica que tiene la enfermedad, ¿cuál es la probabilidad condicional de que tenga, en realidad, la enfermedad? ¿La respuesta lo sorprende? ¿Se considera confiable esta prueba de diagnóstico?

2.126 **Ejercicio Applet** Consulte el Ejercicio 2.125. La probabilidad de que la prueba detecte la enfermedad dado que el paciente tiene la enfermedad se denomina *sensibilidad* de la prueba. La *especificidad* de la prueba es la probabilidad de que la prueba no indique enfermedad dado que el paciente no tiene enfermedades. El *valor de predicción positivo* de la prueba es la probabilidad de que el paciente tenga la enfermedad dado que la prueba indica que la enfermedad está presente. En el Ejercicio 2.125, la enfermedad en cuestión era relativamente rara, presentándose con una probabilidad .01 y la prueba descrita tiene sensibilidad = especificidad = .90 y valor de predicción positivo = .0833.

a En un esfuerzo por aumentar el valor de predicción positivo de la prueba, la sensibilidad se aumentó a .95 y la especificidad permaneció en .90, ¿cuál es el valor de predicción positivo de la prueba "mejorada"?

b Todavía no satisfechos con el valor de predicción positivo del procedimiento, la sensibilidad de la prueba se aumentó a .999. ¿Cuál es el valor de predicción positivo de la prueba modificada (ahora dos veces) si la especificidad permanece en .90?

c Vea con cuidado los diversos números que se emplearon para calcular el valor de predicción positivo de las pruebas. ¿Por qué son tan pequeños todos los valores de predicción positivos? [*Sugerencia:* compare el tamaño del numerador y el denominador empleados en el quebrado que da el valor del valor de predicción positivo. ¿Por qué es (relativamente) tan grande el denominador?]

d La proporción de individuos con la enfermedad no está sujeta a nuestro control. Si la sensibilidad de la prueba es .90, ¿es posible que el valor de predicción positivo de la prueba se pueda aumentar a un valor arriba de .5? ¿Cómo? [*Sugerencia:* considere mejorar la especificidad de la prueba.]

e Con base en los resultados de sus cálculos en las partes previas, si la enfermedad en cuestión es relativamente rara, ¿cómo puede aumentarse significativamente el valor de predicción positivo de una prueba de diagnóstico?

2.127 **Ejercicio Applet** Consulte los Ejercicios 2.125 y 2.126. Suponga ahora que la enfermedad no es particularmente rara y que ocurre con probabilidad .4.

a Si, como en el Ejercicio 2.125, una prueba tiene sensibilidad = especificidad = .90, ¿cuál es el valor de predicción positivo de la prueba?

b ¿Por qué el valor de predicción positivo de la prueba es mucho más alto que el valor obtenido en el Ejercicio 2.125? [*Sugerencia:* compare la magnitud del numerador y el denominador empleados en la fracción que da el valor del valor de predicción positivo.]

c Si la especificidad de la prueba permanece en .90, ¿la sensibilidad de la prueba puede ajustarse para obtener un valor de predicción positivo arriba de .87?

d Si la sensibilidad permanece en .90, ¿la especificidad puede ajustarse para obtener un valor de predicción positivo arriba de .95? ¿Cómo?

e Los inventores de una prueba de diagnóstico desean que la prueba tenga un alto valor de predicción positivo. Con base en los cálculos del lector en partes previas de este problema y en el Ejercicio 2.126, ¿el valor de la especificidad es más o menos crítico cuando se desarrolle una prueba para una enfermedad más rara?

2.128 Use el Teorema 2.8, la ley de probabilidad total, para demostrar lo siguiente:

a Si $P(A|B) = P(A|\overline{B})$, entonces A y B son independientes.

b Si $P(A|C) > P(B|C)$ y $P(A|\overline{C}) > P(B|\overline{C})$, entonces $P(A) > P(B)$.

2.129 Se observa que hombres y mujeres reaccionan de modo diferente a un conjunto determinado de circunstancias; se sabe que 70% de las mujeres reaccionan positivamente a estas circunstancias mientras que de este mismo modo reaccionan sólo 40% de los hombres. Un grupo de 20 personas, 15 mujeres y 5 hombres, se sometió a estas circunstancias y a los sujetos se les pidió describieran sus reacciones en un cuestionario escrito. Una respuesta escogida al azar de las 20 fue negativa. ¿Cuál es la probabilidad de que haya sido de un hombre?

2.130 Un estudio de residentes de Georgia sugiere que quienes trabajaron en astilleros durante la Segunda Guerra Mundial fueron sometidos a un riesgo considerablemente más alto de cáncer pulmonar (*Wall Street Journal*, 21 de septiembre de 1978).[3] Se encontró que alrededor de 22% de quienes tuvieron cáncer pulmonar trabajaron en un tiempo en un astillero. En contraste, sólo 14% de quienes no tuvieron cáncer pulmonar trabajaron en un astillero. Suponga que la proporción de todos los nativos de Georgia que vivieron durante la Segunda Guerra Mundial que han contraído o contraerán cáncer pulmonar es .04%. Encuentre el porcentaje de georgianos que vivieron durante el mismo periodo que contraerán (o ya han contraído) cáncer pulmonar, dado que en algún tiempo han trabajado en un astillero.

2.131 La *diferencia simétrica* entre dos eventos A y B es el conjunto de todos los puntos muestrales que están en *exactamente uno* de los conjuntos y con frecuencia se denota como $A \triangle B$. Observe que $A \triangle B = (A \cap \overline{B}) \cup (\overline{A} \cap B)$. Demuestre que $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$.

2.132 Un avión está extraviado y se presume que tiene igual probabilidad de caer en cualquiera de tres regiones. Si el avión cae realmente en la región i, denote con $1 - \alpha_i$ la probabilidad de que el avión sea hallado en una búsqueda en la i-ésima región, $i = 1, 2, 3$. ¿Cuál es la probabilidad condicional de que el avión se encuentre en la

a región 1, dado que la búsqueda en la región 1 fue infructuosa?

b región 2, dado que la búsqueda en la región 1 fue infructuosa?

c región 3, dado que la búsqueda en la región 1 fue infructuosa?

2.133 Un estudiante contesta una pregunta de opción múltiple de un examen que ofrece cuatro posibles respuestas. Suponga que la probabilidad de que el estudiante conozca la respuesta a la pregunta es .8 y la probabilidad de que el estudiante adivine es .2. Suponga que si el estudiante adivina, la probabilidad de

3. *Fuente: Wall Street Journal*, © Dow Jones & Company, Inc. 1981. Todos los derechos reservados a nivel mundial.

que seleccione la respuesta correcta es .25. Si el estudiante contesta correctamente una pregunta, ¿cuál es la probabilidad de que realmente conozca la respuesta correcta?

2.134 Hay dos métodos, A y B, para enseñar cierta habilidad industrial. El porcentaje de no aprobados es 20% para A y 10% para B, pero B es más costoso y por tanto se usa sólo 30% del tiempo. (A es empleado el otro 70%.) A una trabajadora se le enseñó la habilidad por uno de los dos métodos pero no la aprendió correctamente. ¿Cuál es la probabilidad de que se le haya enseñado por el método A?

2.135 De los viajeros que llegan a un pequeño aeropuerto, 60% vuelan en líneas aéreas importantes, 30% en aviones de propiedad privada y el resto en aviones comerciales que no pertenecen a una línea aérea importante. De quienes viajan en líneas aéreas importantes, 50% viajan por negocios en tanto que 60% de quienes llegan en aviones privados y 90% de quienes llegan en otros aviones comerciales viajan por negocios. Suponga que seleccionamos al azar una persona que llega a este aeropuerto. ¿Cuál es la probabilidad de que la persona

a viaje por negocios?,

b viaje por negocio en un avión privado?,

c llegue en un avión privado, dado que la persona viaja por negocios?,

d viaja por negocio, dado que vuela en un avión comercial?

2.136 Un director de personal tiene dos listas de solicitantes para trabajos. La lista 1 contiene los nombres de cinco mujeres y dos hombres, mientras que la lista 2 contiene los nombres de dos mujeres y seis hombres. Un nombre se selecciona al azar de la lista 1 y se agrega a la lista 2. A continuación se selecciona al azar un nombre de la lista 2 aumentada. Dado que el nombre seleccionado es de un hombre, ¿cuál es la probabilidad de que un nombre de mujer se haya seleccionado originalmente de la lista 1?

2.137 Cinco tazones idénticos están marcados 1, 2, 3, 4 y 5. El tazón i contiene i bolas blancas y $5 - i$ bolas negras, con $i = 1, 2, \ldots, 5$. Un tazón se selecciona al azar y dos bolas se seleccionan al azar (sin restitución) de él.

a ¿Cuál es la probabilidad de que ambas bolas seleccionadas sean blancas?

b Dado que ambas bolas seleccionadas son blancas, ¿cuál es la probabilidad de que el tazón 3 haya sido seleccionado?

***2.138** A continuación está una descripción del juego de *craps.* Un jugador tira dos dados y calcula el total de puntos de sus caras superiores. Si el primer tiro del jugador es un 7 o un 11, el jugador gana el juego. Si el primer tiro es un 2, 3 o 12, el jugador pierde el juego. Si el jugador tira cualquier otra cosa (4, 5, 6, 8, 9 o 10) en el primer tiro, el valor será el *punto* del jugador. Si el jugador no gana ni pierde en el primer tiro, lanza los dados en forma repetida hasta que obtenga su punto o un 7. Gana si tira su punto antes de tirar un 7 y pierde si tira un 7 antes de su punto. ¿Cuál es la probabilidad de que el jugador gane un juego de craps? [*Sugerencia:* recuerde el Ejercicio 2.119.]

2.11 Eventos numéricos y variables aleatorias

Los eventos de mayor interés para el científico, ingeniero u hombre de negocios son los identificados por números, llamados *eventos numéricos.* El médico investigador está interesado en el evento de que diez de cada diez pacientes tratados sobreviva a una enfermedad; el hombre de negocios está interesado en el evento de que sus ventas en el año próximo lleguen a 5 millones de dólares. Denote con Y una variable a ser medida en un experimento. Como el valor de Y varía dependiendo del resultado del experimento, se denomina *variable aleatoria.*

A cada punto del espacio muestral le asignaremos un número real que denote el valor de la variable Y. El valor asignado a Y varía de un punto muestral a otro, pero a algunos puntos se

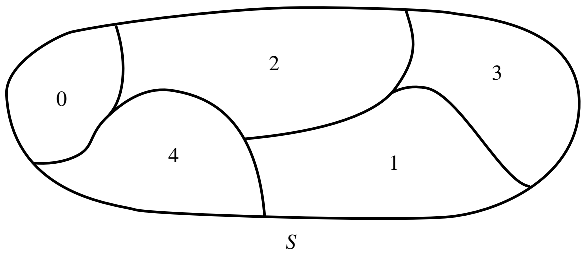


FIGURA **2.14** División de S en subconjuntos que definen los eventos $Y = 0, 1, 2, 3$ y 4

les puede asignar el mismo valor numérico. Entonces, hemos definido una variable que es una función de los puntos muestrales en S y {todos los puntos muestrales donde $Y = a$} es el evento numérico asignado al número a. De hecho, el espacio muestral S se puede dividir en subconjuntos para que a todos los puntos dentro de un subconjunto se les asigne el mismo valor de Y. Estos subconjuntos son mutuamente excluyentes porque a ningún punto se le asignan dos valores numéricos diferentes. La división de S está simbólicamente indicada en la Figura 2.14 para una variable aleatoria que puede tomar valores 0, 1, 2, 3 y 4.

DEFINICIÓN **2.12**

Una *variable aleatoria* es una función de valor real para la cual el dominio es un espacio muestral.

EJEMPLO **2.24** Definir un experimento como lanzar dos monedas al aire y observar los resultados. Sea Y igual al número de caras obtenido. Identifique los puntos muestrales en S, asigne un valor de Y a cada punto muestral e identifique los puntos muestrales asociados con cada valor de la variable aleatoria Y.

Solución Represente con H y T cara y cruz, respectivamente, e identifique con un par ordenado de símbolos al resultado de las monedas primera y segunda. (Así, HT implica una cara en la primera moneda y una cruz en la segunda.) Entonces los cuatro puntos muestrales en S son E_1: HH, E_2: HT, E_3: TH y E_4: TT. Los valores de Y asignados a los puntos muestrales dependen del número de caras asociado con cada punto. Para E_1: HH, se observaron dos caras, y a E_1 se le asigna el valor $Y = 2$. Del mismo modo, asignamos los valores $Y = 1$ a E_2 y E_3, y $Y = 0$ a E_4. Resumiendo, la variable aleatoria Y puede tomar tres valores, $Y = 0$, 1 y 2, que son eventos definidos por conjuntos de puntos muestrales:

$$\{Y = 0\} = \{E_4\}, \qquad \{Y = 1\} = \{E_2,\ E_3\}, \qquad \{Y = 2\} = \{E_1\}. \qquad \blacksquare$$

Denote con y un valor observado de la variable aleatoria Y. Entonces $P(Y = y)$ es la suma de las probabilidades de los puntos muestrales a los que se asigna el valor y.

EJEMPLO **2.25** Calcule las probabilidades para cada valor de Y del Ejemplo 2.24.

Solución El evento $\{Y = 0\}$ resulta sólo del punto muestral E_4. Si las monedas están balanceadas, los puntos muestrales son igualmente probables; por tanto,

$$P(Y = 0) = P(E_4) = 1/4.$$

Del mismo modo,

$$P(Y = 1) = P(E_2) + P(E_3) = 1/2 \quad \text{y} \quad P(Y = 2) = P(E_1) = 1/4. \qquad \blacksquare$$

En los dos capítulos siguientes veremos un examen más detallado de variables aleatorias.

Ejercicios

2.139 Consulte el Ejercicio 2.112. Con la variable aleatoria Y represente al número de aparatos de radar que detectan un avión particular. Calcule las probabilidades asociadas con cada valor de Y.

2.140 Consulte el Ejercicio 2.120. Con la variable aleatoria Y represente el número de refrigeradores defectuosos hallados después de probar tres de ellos. Calcule las probabilidades para cada valor de Y.

2.141 Consulte de nuevo el Ejercicio 2.120. Con la variable aleatoria Y represente el número de la prueba en la que se identifica el último refrigerador defectuoso. Calcule las probabilidades para cada valor de Y.

2.142 Un proyectil puede aterrizar en cualquiera de cuatro posiciones, A, B, C y D, con igual probabilidad. El proyectil se usa dos veces y la posición se anota en cada vez. Con la variable aleatoria Y denote el número de posiciones en las que el proyectil *no* aterrizó. Calcule las probabilidades para cada valor de Y.

2.12 Muestreo aleatorio

Como tema final en este capítulo pasamos de la teoría a la aplicación y examinamos la naturaleza de experimentos realizados en estadística. Un experimento estadístico involucra la observación de una muestra seleccionada de un conjunto más grande de datos, existente o conceptual, llamado *población*. Las mediciones de la muestra, vistas como observaciones de los valores de una o más variables aleatorias, se emplean entonces para hacer una inferencia acerca de las características de la población objetivo.

¿Cómo se hacen estas inferencias? Una respuesta exacta a esta pregunta se pospone para más adelante, pero una observación general se deduce de nuestra exposición en la Sección 2.2. Ahí aprendimos que la probabilidad de la muestra observada desempeña un importante papel para hacer una inferencia y evaluar su credibilidad.

Sin extendernos sobre el punto, es claro que el método de muestreo afectará la probabilidad del resultado de una muestra particular. Por ejemplo, supongamos que una población ficticia contiene sólo $N = 5$ elementos, de los cuales pensamos tomar una muestra de tamaño $n = 2$.

Se podrían revolver muy bien los elementos y seleccionar dos en forma tal que todos los pares de elementos tengan igual probabilidad de selección. Un segundo procedimiento de muestreo podría hacer necesario seleccionar un solo elemento, restituirlo en la población y después sacar de nuevo un solo elemento. Los dos métodos de selección de muestra se denominan *muestreo sin* y *con reemplazo*, respectivamente.

Si los todos los elementos de la población $N = 5$ son claramente diferentes, la probabilidad de sacar un par específico, cuando se haga muestreo sin restitución, es 1/10. La probabilidad de sacar el mismo par específico, cuando se haga muestreo con restitución, es 2/25. Con toda facilidad se pueden verificar estos resultados.

Se debe hacer notar que el método de muestreo, conocido como *diseño de un experimento*, afecta la cantidad de información en una muestra y la probabilidad de observar un resultado específico de muestra. Por tanto, todo procedimiento muestral debe describirse con claridad si deseamos hacer inferencias válidas de muestra a población.

El estudio del diseño de experimentos, los diversos tipos de diseños junto con sus propiedades, es en sí mismo un curso. En consecuencia, en esta primera etapa de estudio presentamos sólo el procedimiento de muestreo más sencillo, el *muestreo aleatorio simple.* La noción de muestreo aleatorio simple será necesaria en subsecuentes exposiciones de las probabilidades asociadas con variables aleatorias e inyectará algún realismo a nuestro examen de estadística. Eso es porque el muestreo aleatorio simple se emplea frecuentemente en la práctica. Definamos ahora el término *muestreo aleatorio.*

DEFINICIÓN 2.13 Represente con N y n los números de elementos en la población y la muestra, respectivamente. Si el muestreo se realiza en forma tal que cada una de las $\binom{N}{n}$ muestras tiene igual probabilidad de ser seleccionada, se dice que el muestreo es aleatorio y que el resultado es una *muestra aleatoria.*

El muestreo aleatorio perfecto es difícil de alcanzar en la práctica. Si la población no es demasiado grande, podríamos escribir cada uno de los N números en fichas, revolverlas todas y seleccionar una muestra de n fichas. Los números de las fichas especificarían las mediciones que aparezcan en la muestra.

En computadora se han elaborado tablas de números aleatorios para agilizar la selección de muestras aleatorias. Un ejemplo de esas tablas es la Tabla 12, Apéndice 3. Una tabla de número aleatorio es un conjunto de enteros $(0, 1, \ldots, 9)$ generado de modo que, a la larga, la tabla contendrá los diez enteros en proporciones aproximadamente iguales, sin tendencias en las formas en las que los dígitos se generan. En consecuencia, si un dígito se selecciona de un punto aleatorio en la tabla, es igualmente probable que sea cualquiera de los dígitos del 0 al 9.

Escoger números de la tabla es análogo a sacar fichas numeradas de una pila revuelta, como ya dijimos. Suponga que deseamos una muestra aleatoria de tres personas a seleccionarse de una población de siete personas. Podríamos numerar las personas del 1 al 7, poner los números en fichas, revolver muy bien las fichas y luego sacar tres de ellas. De una forma análoga, podrían poner la punta de un lápiz en un punto inicial aleatorio de la Tabla 12, Apéndice 3. Suponga que la punta cae en la línea 15 de la columna 9 y que decidimos usar el dígito de la extrema derecha del grupo de cinco, que es un 5 en este caso. Este proceso es como sacar una ficha con el número 5. Ahora proseguiríamos en cualquier dirección para obtener los números restantes de la muestra. Si decidimos continuar hacia abajo de la página, el siguiente número (inmediatamente abajo del 5) es un 2. Entonces nuestra segunda persona muestreada sería el número 2. Continuando, llegamos a un 8 pero hay sólo siete elementos en la población. Entonces el 8 se pasa por alto y continuamos bajando por la columna; aparecen entonces dos números 5 pero ambos deben ser ignorados porque la persona 5 ya ha sido seleccionada. (La ficha número 5 se ha retirado del montón.) Por último, llegamos a un 1 y nuestra muestra de tres está completa con personas numeradas 5, 2 y 1.

Cualquier punto inicial se puede usar en una tabla de números aleatorios y podemos continuar en cualquier dirección desde el punto inicial. No obstante, si se ha de usar más de una muestra en cualquier problema, cada una debe tener un punto inicial único.

En muchas situaciones la población es conceptual, como en una observación hecha durante un experimento de laboratorio. Aquí la población se visualiza como las infinitamente numerosas mediciones que se obtendrían si el experimento fuera a repetirse una y otra vez. Si deseamos una muestra de $n = 10$ mediciones de esta población, repetiríamos el experimento 10 veces y esperaríamos que los resultados representen, con un grado razonable de aproximación, una muestra aleatoria.

Aun cuando el propósito principal de esta exposición fue aclarar el significado de una muestra aleatoria, nos gustaría mencionar que algunas técnicas de muestreo son sólo parcialmente aleatorias. Por ejemplo, si deseamos determinar la preferencia electoral de la nación en una elección presidencial, es probable que no escojamos una muestra aleatoria de la población de votantes. Por pura casualidad, todos los votantes que aparezcan en la muestra podrían ser sacados de una sola ciudad, San Francisco, por ejemplo, que no podría ser representativa de la población de todos los votantes de Estados Unidos. Preferiríamos una selección aleatoria de votantes provenientes de distritos políticos más pequeños, quizá estados, asignando un número específico a cada estado. La información de las submuestras seleccionadas al azar sacadas de los estados respectivos se combinaría para formar una predicción respecto a toda la población de votantes del país. En general, deseamos seleccionar una muestra para obtener una cantidad específica de información a un costo mínimo.

2.13 Resumen

Este capítulo se ha ocupado de dar un modelo para la repetición de un experimento y, en consecuencia, un modelo para las distribuciones de frecuencia de población del Capítulo 1. La adquisición de una distribución de probabilidad es el primer paso para formar una teoría que modele la realidad y para desarrollar la maquinaria para hacer inferencias.

Un experimento se definió como el proceso de hacer una observación. Los conceptos de evento, evento simple, espacio muestral y los axiomas de probabilidad han dado un modelo probabilístico para calcular la probabilidad de un evento. Los eventos numéricos y la definición de una variable aleatoria se introdujeron en la Sección 2.11.

Inherente al modelo es el método de punto muestral para calcular la probabilidad de un evento (Sección 2.5). Las reglas de conteo útiles para aplicar el método de punto muestral se estudiaron en la Sección 2.6. El concepto de probabilidad condicional, las operaciones de álgebra de conjuntos, así como las leyes de probabilidad ponen el escenario para el método de composición de evento para calcular la probabilidad de un evento (Sección 2.9).

¿De qué valor es la teoría de probabilidad? Proporciona la teoría y las herramientas para calcular las probabilidades de eventos numéricos y por tanto las distribuciones de probabilidad para las variables aleatorias que se estudiarán en el Capítulo 3. Los eventos numéricos de interés para nosotros aparecen en una muestra y vamos a calcular la probabilidad de una muestra observada para hacer una inferencia acerca de la población objetivo. La probabilidad proporciona la base y las herramientas para la inferencia estadística, objetivo de la estadística.

Bibliografía y lecturas adicionales

Cramer, H. 1973. The *Elements of Probability Theory and Some of Its Applications*, 2d ed.Huntington, N. Y.: Krieger.

Feller, W. 1968. *An Introductiont to Probability Theory and Its Applications*, 3d ed., vol.1. NewYork: Wiley.

———. 1971. *An Introductiont to Probability Theory and Its Applications*, 2d ed., vol.2. New York: Wiley.

Meyer, P. L. 1970. *Introductory Probability and Statistical Applications*, 2d ed. Reading, Mass.: Addison-Wesley.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Riordan, J. 2002. *Introduction to Combinatorial Analysis. Mineola*, N. Y.: Dover Publications.

Ejercicios complementarios

2.143 Demuestre que el Teorema 2.7 se cumple para probabilidades *condicionales.* Esto es, si $P(B) > 0$, entonces $P(A|B) = 1 - P(\overline{A}|B)$.

2.144 Suponga que S contiene cuatro puntos muestrales, E_1, E_2, E_3 y E_4.

a Haga una lista de todos los posibles eventos en S (incluyendo el evento nulo).

b En el Ejercicio 2.68(d), se demostró que $\sum_{i=1}^{n} \binom{n}{i} = 2^n$. Use este resultado para dar el número total de eventos en S.

c Sean A y B los eventos $\{E_1, E_2, E_3\}$ y $\{E_2, E_4\}$, respectivamente. Escriba los puntos muestrales en los siguientes eventos: $A \cup B$, $A \cap B$, $\overline{A} \cap \overline{B}$ y $\overline{A} \cup B$.

2.145 Un paciente que se somete a un examen físico anual debe pasar por 18 verificaciones o pruebas. La secuencia en la que las pruebas se realizan es importante porque el tiempo perdido entre pruebas va a variar dependiendo de la secuencia. Si una experta en eficiencia fuera a estudiar las secuencias para hallar la que requirió menor tiempo, ¿cuántas secuencias estarían incluidas en su estudio si todas las posibles secuencias fueran admisibles?

2.146 Se sacan cinco cartas de una baraja estándar de 52 cartas. ¿Cuál es la probabilidad de que las cinco sean del mismo “palo”?

2.147 Consulte el Ejercicio 2.146. Un jugador ha recibido cinco cartas: dos ases, un rey, un cinco y un 9. Él desecha el 5 y el 9 y recibe dos cartas más. ¿Cuál es la probabilidad de que termine con un *full*?

2.148 Una tolva contiene tres componentes provenientes del proveedor A, cuatro del proveedor B y cinco del proveedor C. Si cuatro de los componentes se seleccionan al azar para probarlos, ¿cuál es la probabilidad de que cada proveedor tenga al menos un componente probado?

2.149 Un grupo grande de personas se examinará para ver si presentan dos síntomas comunes de cierta enfermedad. Se piensa que 20% de las personas tienen sólo el síntoma A, 30% sólo el síntoma B, 10% tienen ambos síntomas y el resto no tiene ningún síntoma. Para una persona escogida al azar de este grupo, encuentre las siguientes probabilidades:

a La persona no tiene ningún síntoma.

b La persona tiene al menos un síntoma.

c La persona tiene ambos síntomas, dado que tiene el síntoma B.

2.150 Consulte el Ejercicio 2.149. Con la variable aleatoria Y represente el número de síntomas que tiene una persona escogida al azar del grupo. Calcule las probabilidades asociadas con cada uno de los valores de Y.

***2.151** **Un modelo para la Serie Mundial** Dos equipos A y B juegan una serie de partidos hasta que un equipo gane cuatro de ellos. Suponemos que los partidos se juegan de manera independiente y que la probabilidad de que A gane cualquier juego es p. ¿Cuál es la probabilidad de que la serie dure exactamente cinco partidos?

2.152 Sabemos lo siguiente acerca de un método colorimétrico que se utiliza para probar la existencia de nitratos en el agua de lagos. Si los especímenes del agua contienen nitratos, una solución que se aplica al agua hará que el espécimen se vuelva rojo 95% de las veces. Cuando se usa en especímenes de agua sin nitratos, la solución hace que el agua se vuelva roja 10% de las veces (porque los elementos químicos que no sean nitratos se presentan ocasionalmente y también reaccionan al agente). La experiencia del pasado en un laboratorio indica que los nitratos están contenidos en 30% de los especímenes de agua que se envían al laboratorio para someterlos a pruebas. Si un espécimen de agua se selecciona al azar

a de entre los enviados al laboratorio, ¿cuál es la probabilidad de que se vuelva rojo cuando se pruebe?,

b y si se vuelve rojo cuando se prueba, ¿cuál es la probabilidad de que en realidad contenga nitratos?

2.153 Algunos casos de historias clínicas indican que diferentes enfermedades producen síntomas idénticos. Suponga que un conjunto particular de síntomas, denotado como H, se presenta sólo con cualquiera de tres enfermedades, I_1, I_2 o I_3. Suponga que la presentación simultánea de más de una de estas enfermedades es imposible y que

$$P(I_1) = .01, \qquad P(I_2) = .005, \qquad P(I_3) = .02.$$

Las probabilidades de desarrollar el conjunto de síntomas H, dada cada una de estas enfermedades, se sabe que son

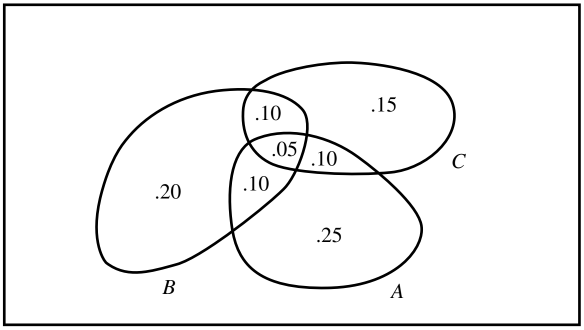
$$P(H|I_1) = .90, \qquad P(H|I_2) = .95, \qquad P(H|I_3) = .75.$$

Suponiendo que una persona enferma presenta los síntomas, H, ¿cuál es la probabilidad de que la persona tenga la enfermedad I_1?

2.154 **a** Un cajón contiene $n = 5$ pares de calcetines diferentes y distinguibles (un total de diez calcetines). Si una persona (quizá en la oscuridad) selecciona al azar cuatro calcetines, ¿cuál es la probabilidad de que no haya un par igual en la muestra?

***b** Un cajón contiene n pares de calcetines diferentes y distinguibles (un total de $2n$ calcetines). Una persona selecciona al azar $2r$ de los calcetines, donde $2r < n$. En términos de n y r, ¿cuál es la probabilidad de que no haya un par igual en la muestra?

2.155 Un grupo de hombres comparten las características de estar casados (A), tener un título universitario (B) y ser ciudadanos de un estado especificado (C), de acuerdo con las fracciones dadas en el siguiente diagrama de Venn. Esto es, 5% de los hombres tienen las tres características, mientras que 20% tienen educación universitaria pero no están casados y no son ciudadanos del estado especificado. Un hombre se selecciona al azar de este grupo.



Encuentre la probabilidad de que él

a sea casado,

b tenga un título universitario y sea casado,

c no sea del estado especificado pero es casado y tiene un título universitario,

d no está casado ni tiene título universitario, dado que es del estado especificado.

2.156 La siguiente tabla indica las muertes accidentales por edad y ciertos tipos específicos para Estados Unidos en 2002.

a Se supo que una persona seleccionada al azar de Estados Unidos tuvo una muerte accidental en 2002. Encuentre la probabilidad de que

i haya tenido más de 15 años,

ii la causa de su muerte haya sido un accidente automovilístico,

iii la causa de su fuerte haya sido un accidente automovilístico, dado que la persona tenía entre 15 y 24 años de edad,

iv La causa de su muerte haya sido por ahogamiento, dado que no fue un accidente automovilístico y que la persona tenía 34 años o menos.

b De estas cifras, ¿se puede determinar la probabilidad de que una persona seleccionada al azar de la población de Estados Unidos haya tenido un accidente automovilístico mortal en 2002?

	Tipo de accidente			
Edad	Todos tipos	Accidente automovilístico	Caída	Ahogamiento
Menos de 5	2,707	819	44	568
5–14	2,979	1,772	37	375
15–24	14,113	10,560	237	646
25–34	11,769	6,884	303	419
35–44	15,413	6,927	608	480
45–54	12,278	5,361	871	354
55–64	7,505	3,506	949	217
65–74	7,698	3,038	1,660	179
75 o más	23,438	4,487	8,613	244
Total	97,900	43,354	13,322	3,482

Fuente: compilado de *National Vital Statistics Report* 50, núm.15, 2002.

2.157 Un estudio de los residentes de una región mostró que 20% eran fumadores. La probabilidad de muerte debida a cáncer de pulmón si la persona fuma es diez veces mayor que la probabilidad de muerte si la persona no fuma. Si la probabilidad de muerte debida a cáncer de pulmón en la región es .006, ¿cuál es la probabilidad de muerte debida a cáncer de pulmón dado que la persona es fumadora?

2.158 Un tazón contiene w bolas blancas y b bolas negras. Una bola se selecciona al azar del tazón, se anota su color y se devuelve al tazón junto con n bolas adicionales del mismo color. Otra bola individual se selecciona al azar del tazón (que ahora contiene $w + b + n$ bolas) y se observa que la bola es negra. Demuestre que la probabilidad (condicional) de que la primera bola seleccionada sea blanca es $\dfrac{w}{w + b + n}$.

2.159 Parece obvio que $P(\emptyset) = 0$. Demuestre que este resultado se deduce de los axiomas de la Definición 2.6.

2.160 Una máquina para producir un nuevo componente electrónico experimental genera piezas defectuosas de vez en cuando en una forma aleatoria. El ingeniero supervisor para una máquina particular ha observado que piezas defectuosas parecen estar agrupándose (es decir apareciendo de un modo no aleatorio), lo cual sugiere un mal funcionamiento en alguna parte de la máquina. Una prueba para la aleatoriedad se basa en el número de *lotes* de piezas defectuosas y no defectuosas (un lote es una secuencia ininterrumpida de piezas defectuosas y no defectuosas). Cuanto más pequeño es el número de lotes, mayor será la cantidad de evidencia que indica la no aleatoriedad. De 12 componentes sacados de la máquina, los primeros 10 no fueron defectuosos y los últimos 2 fueron defectuosos ($N\,N\,N\,N\,N\,N\,N\,N\,N\,N\,D\,D$). Suponga que no hay aleatoriedad. ¿Cuál es la probabilidad de observar

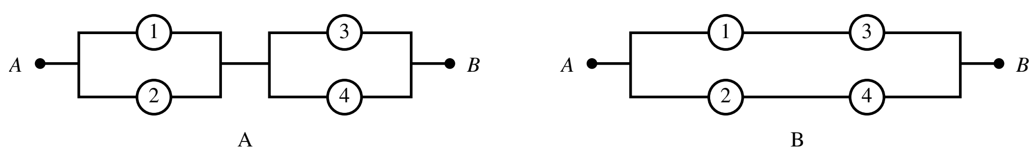
a este arreglo (resultante en dos lotes) dado que 10 de los 12 componentes no son defectuosos?

b dos lotes?

2.161 Consulte el Ejercicio 2.160. ¿Cuál es la probabilidad de que el número de lotes, R, sea menor o igual a 3?

2.162 Suponga que hay nueve espacios de estacionamiento juntos en un lote. Nueve autos necesitan ser estacionados por un empleado. Tres de los autos son costosos autos deportivos, tres son grandes autos de fabricación nacional y tres son compactos importados. Suponiendo que el empleado estaciona los autos al azar, ¿cuál es la probabilidad de que los tres costosos autos deportivos queden estacionados uno junto al otro?

2.163 Los relevadores empleados en la construcción de circuitos eléctricos funcionan de manera adecuada con probabilidad .9. Suponiendo que el circuito opera independientemente, ¿cuál de los siguientes diseños de circuito da la más alta probabilidad de que circule corriente cuando los relevadores se activen?



2.164 Consulte el Ejercicio 2.163 y considere el circuito A. Si sabemos que la corriente está circulando, ¿cuál es la probabilidad de que los interruptores 1 y 4 estén funcionando correctamente?

2.165 Consulte el Ejercicio 2.163 y considere el circuito B. Si sabemos que la corriente está circulando, ¿cuál es la probabilidad de que los interruptores 1 y 4 estén funcionando correctamente?

2.166 Ocho neumáticos de marcas diferentes se clasifican del 1 al 8 (del mejor al peor) de acuerdo con las millas recorridas. Si cuatro de estos neumáticos son seleccionados al azar por un cliente, encuentre la probabilidad de que el mejor de entre los seleccionados por el cliente esté en realidad clasificado tercero entre los ocho originales.

2.167 Consulte el Ejercicio 2.166. Denote con Y la clasificación real de calidad del mejor neumático seleccionado por el cliente. En el Ejercicio 2.166 se calculó $P(Y = 3)$. Dé los posibles valores de Y y las probabilidades asociadas con todos estos valores.

2.168 Al igual que en los Ejercicios 2.166 y 2.167, ocho neumáticos de diferentes marcas fueron clasificados del 1 al 8 (del mejor al peor) de acuerdo con las millas recorridas.

a Si cuatro de estos neumáticos son seleccionados al azar por un cliente, ¿cuál es la probabilidad de que el mejor neumático seleccionado sea clasificado 3 y el peor sea clasificado 7?

b En el inciso a se calculó la probabilidad de que el mejor neumático seleccionado esté clasificado 3 y el peor esté clasificado 7. Si ése es el caso, la *amplitud* de las clasificaciones, R = clasificación más grande − clasificación más pequeña $= 7 - 3 = 4$. ¿Cuál es $P(R = 4)$?

c Dé todos los posibles valores para R y las probabilidades asociadas con todos estos valores.

***2.169** Tres bebedores de cerveza (I, II y III) van a clasificar cuatro marcas diferentes de cerveza (A, B, C y D) en una prueba a ciegas. Cada bebedor clasifica las cuatro cervezas como 1 (para la cerveza que más le gustó), 2 (para la siguiente mejor), 3 o 4.

a Con todo cuidado describa un espacio muestral para este experimento (observe que necesitamos especificar la clasificación de las cuatro cervezas para los tres bebedores). ¿Cuántos puntos muestrales hay en el espacio muestral?

b Suponga que los bebedores no pueden discriminar entre las cervezas, de modo que cada asignación de lugares a las cervezas es igualmente probable. Después de que las cervezas han sido clasificadas por los bebedores, se suman los lugares de cada marca de cerveza. ¿Cuál es la probabilidad de que alguna cerveza reciba una clasificación total de 4 o menos?

2.170 Tres nombres se van a seleccionar de una lista de siete para una encuesta de opinión pública. Encuentre la probabilidad de que el primer nombre de la lista sea seleccionado para la encuesta.

2.171 Un caso del servicio de noticias AP, impreso en el *Gainesville Sun* el 20 de mayo de 1979, indica lo siguiente con respecto a los restos del *Skylab* que golpearon a alguien en tierra: "Las probabilidades son 1 en 150 de que una pieza del Skylab golpee a alguien. Pero 4 mil millones de personas. . . viven en la zona en la que podrían caer piezas. Entonces, las probabilidades de que cualquier persona sea golpeada son una en 150 por 4 mil millones o sea una en 600 mil millones". ¿Ve el lector algunas imprecisiones en este razonamiento?

2.172 Sean A y B dos eventos. ¿Cuáles de los siguientes planteamientos, en general, son falsos?

a $P(A|B) + P(\overline{A}|\overline{B}) = 1.$

b $P(A|B) + P(A|\overline{B}) = 1.$

c $P(A|B) + P(\overline{A}|B) = 1.$

2.173 A medida que salen piezas del final de una línea de producción, un inspector selecciona cuáles han de someterse a una inspección completa. Diez por ciento de todas las piezas producidas son defectuosas. Sesenta por ciento de todas las piezas defectuosas pasan por una inspección completa y 20% de todas las piezas en buenas condiciones pasan por una inspección completa. Dado que una pieza se inspecciona por completo, ¿cuál es la probabilidad de que sea defectuosa?

2.174 Muchas escuelas públicas están poniendo en práctica una regla de "pasa o no pasa" para atletas. De acuerdo con este sistema, un estudiante que no apruebe un curso es descalificado para participar en actividades extracurriculares durante el siguiente periodo de calificación. Suponga que la probabilidad es .15 de que un atleta que no ha sido descalificado previamente sea descalificado en el siguiente periodo. Para atletas que han sido descalificados previamente, la probabilidad de descalificación en el siguiente periodo es .5. Si 30% de los atletas han sido descalificados en periodos previos. ¿cuál es la probabilidad de que un atleta seleccionado al azar sea descalificado durante el siguiente periodo de calificación?

2.175 Se dice tres eventos, A, B y C que son mutuamente independientes si

$$P(A \cap B) = P(A) \times P(B), \qquad P(B \cap C) = P(B) \times P(C),$$
$$P(A \cap C) = P(A) \times P(C), \qquad P(A \cap B \cap C) = P(A) \times P(B) \times P(C).$$

Suponga que una moneda balanceada se lanza al aire dos veces, de manera independiente. Defina los siguientes eventos:

A: aparece una cara en el primer tiro.

B: aparece una cara en el segundo tiro.

C: ambos tiros dan el mismo resultado.

¿A, B y C son mutuamente independientes?

2.176 Consulte el Ejercicio 2.175 y suponga que los eventos A, B y C son mutuamente independientes.

a Demuestre que $(A \cup B)$ y C son independientes.

b Demuestre que A y $(B \cap C)$ son independientes.

2.177 Consulte el Ejercicio 2.90(b) donde un amigo dijo que si hay 1 en 50 probabilidades de lesión en un solo salto, entonces hay un 100% de probabilidad de lesión si un paracaidista salta 50 veces. Suponga que los resultados de saltos repetidos son mutuamente independientes.

a ¿Cuál es la probabilidad de que se completen 50 saltos sin una lesión?

b ¿Cuál es la probabilidad de que ocurra al menos una lesión en 50 saltos?

c ¿Cuál es el número máximo de saltos, n, que el paracaidista puede hacer si la probabilidad es al menos .60 de que los n saltos se completen sin lesión?

***2.178** Suponga que la probabilidad de exposición a la gripe durante una epidemia es .6. La experiencia ha demostrado que un suero tiene 80% de éxito para prevenir que una persona inoculada contraiga la gripe si se expone a ella. Una persona no inoculada enfrenta una probabilidad de .90 de contraer la gripe si se expone a ella. Dos personas, una inoculada y otra no, realizan un trabajo altamente especializado en un negocio. Suponga que no están en el mismo lugar, no están en contacto con las mismas personas y no pueden contagiarse entre sí a la gripe. ¿Cuál es la probabilidad de que al menos una de ellas se enferme?

***2.179** Dos jugadores apuestan \$1 cada uno en los tiros sucesivos al aire de una moneda. Cada uno tiene una banca de \$6. ¿Cuál es la probabilidad de que

a salgan a mano después de seis tiros de la moneda?,

b un jugador, Jones, por ejemplo, gane todo el dinero en el décimo tiro de la moneda?

***2.180** Suponga que las calles de una ciudad son trazadas en una cuadrícula con calles corriendo de norte a sur y de este a oeste. Considere el siguiente esquema para patrullar una zona de 16 manzanas por 16 manzanas. Un oficial empieza a caminar en el crucero del centro de la zona. En la esquina de cada manzana el oficial elige al azar ir al norte, sur, este u oeste. ¿Cuál es la probabilidad de que el oficial

a llegue al límite de la zona de patrullaje después de caminar las primeras 8 manzanas?,

b regrese al punto de partida después de caminar exactamente 4 manzanas?,

***2.181** Suponga que n pelotas que no se pueden distinguir unas de otras se van a acomodar en N cajas que sí se pueden distinguir, de modo que cada arreglo distinguible sea igualmente probable. Si $n \geq N$, demuestre que la probabilidad de que no haya una caja vacía está dada por

$$\frac{\binom{n-1}{N-1}}{\binom{N+n-1}{N-1}}.$$

CAPÍTULO 3

Variables aleatorias discretas y sus distribuciones de probabilidad

3.1 Definición básica

Como ya dijimos en la Sección 2.12, una variable aleatoria es una función de valor real definida sobre un espacio muestral. Por tanto, una variable aleatoria se puede usar para identificar eventos numéricos que son de interés en un experimento. Por ejemplo, el evento de interés en un sondeo de opinión con respecto a las preferencias de votantes no suele ser la persona particular muestreada o el orden en el que se obtuvieron las preferencias, sino $Y =$ el *número* de votantes que están a favor de cierto candidato o tema. El valor observado de esta variable aleatoria debe ser cero o un entero entre 1 y el tamaño muestral. Entonces, esta variable aleatoria puede tomar sólo un número finito de valores con probabilidad diferente de cero. Se dice que una variable aleatoria de este tipo es discreta.

DEFINICIÓN **3.1**

Se dice que una variable aleatoria Y es *discreta* si puede tomar sólo un número finito o contablemente infinito[1] de valores distintos.

Se puede obtener una caracterización menos formidable de variables aleatorias discretas si se consideran algunos ejemplos prácticos. El número de bacterias por unidad de área en el estudio de control de medicamentos sobre el crecimiento de bacterias es una variable aleatoria discreta, como lo es el número de televisores defectuosos en un envío de 100 aparatos. De hecho, las variables aleatorias discretas representan con frecuencia cantidades asociadas con fenómenos reales.

Consideremos ahora la relación entre el material del Capítulo 2 y el de este capítulo. ¿Por qué estudiar la teoría de probabilidad? La respuesta es que se necesita la probabilidad de un evento observado para hacer inferencias acerca de una población. Los eventos de interés son con frecuencia numéricos y corresponden a valores de variables aleatorias discretas. Por tanto, es imperativo que conozcamos las probabilidades de estos eventos numéricos. Como ciertos tipos de variables aleatorias se presentan frecuentemente en la práctica, es útil tener a mano la probabilidad para cada valor de una variable aleatoria. Este conjunto de probabilidades recibe el nombre de *distribución de probabilidad* de la variable aleatoria discreta. Encontraremos que numerosos experimentos muestran características similares y generan variables aleatorias con el mismo tipo de distribución de probabilidad. En consecuencia, el conocimiento de las distribuciones de probabilidad para variables aleatorias asociadas con tipos comunes de experimentos eliminará la necesidad de resolver los mismos problemas de probabilidad una y otra vez.

3.2 La distribución de probabilidad para una variable aleatoria discreta

De modo notacional, usaremos una *letra mayúscula*, por ejemplo Y, para denotar una *variable aleatoria* y una *letra minúscula,* por ejemplo y, para denotar un *valor particular* que puede tomar una variable aleatoria. Por ejemplo, denotemos con Y cualquiera de los seis posibles valores que podrían ser observados en la cara superior cuando se tira un dado. Después de tirar el dado, el número observado será denotado por el símbolo y. Observe que Y es una variable aleatoria, pero el valor específico observado, y, no es aleatorio.

La expresión $(Y = y)$ se puede leer como *el conjunto de todos los puntos en S a los que la variable aleatoria Y asigna el valor y*.

Ahora es significativo hablar de la probabilidad de que Y tome el valor y, denotado por $P(Y = y)$. Al igual que en la Sección 2.11, esta probabilidad está definida como la suma de las probabilidades de puntos muestrales apropiados en S.

1. Recuerde que un conjunto de elementos es contablemente infinito si los elementos del conjunto se pueden poner en correspondencia biunívoca con los enteros positivos.

DEFINICIÓN **3.2**

La probabilidad de que Y tome el valor y, $P(Y = y)$, se define como la *suma de las probabilidades de todos los puntos muestrales en S* a los que se asigna el valor y. A veces denotaremos $P(Y = y)$ por $p(y)$.

Como $p(y)$ es una función que asigna probabilidades a cada valor y de la variable aleatoria Y, a veces recibe el nombre de *función de probabilidad* para Y.

DEFINICIÓN **3.3**

La *distribución de probabilidad* para una variable discreta Y puede ser representada por una fórmula, una tabla o una gráfica que produzca $p(y) = P(Y = y)$ para toda y.

Nótese que $p(y) \geq 0$ para toda y, pero la distribución de probabilidad para una variable aleatoria discreta asigna probabilidades diferentes de cero a sólo un número contable de valores y distintos. Se entiende que cualquier valor y no explícitamente asignado a una posibilidad positiva es tal que $p(y) = 0$. Ilustramos estas ideas con un ejemplo.

EJEMPLO **3.1** Un supervisor en una planta manufacturera tiene tres hombres y tres mujeres trabajando para él y desea escoger dos trabajadores para un trabajo especial. No queriendo mostrar sesgo en su selección, decide seleccionar los dos trabajadores al azar. Denote con Y el número de mujeres en su selección. Encuentre la distribución de probabilidad para Y.

Solución El supervisor puede seleccionar dos trabajadores de entre seis en $\binom{6}{2} = 15$ formas. Entonces, S contiene 15 puntos muestrales que suponemos son igualmente probables porque se empleó muestreo aleatorio. Así, $P(E_i) = 1/15$, para $i = 1, 2, \ldots, 15$. Los valores para Y que tienen probabilidad diferente de cero son 0, 1 y 2. El número de formas de seleccionar $Y = 0$ mujeres es $\binom{3}{0}\binom{3}{2}$ porque el supervisor debe seleccionar cero trabajadores de las tres mujeres y dos de los tres hombres. Entonces, hay $\binom{3}{0}\binom{3}{2} = 1 \cdot 3 = 3$ puntos muestrales en el evento $Y = 0$, y

$$p(0) = P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = \frac{3}{15} = \frac{1}{5}.$$

Del mismo modo,

$$p(1) = P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = \frac{9}{15} = \frac{3}{5},$$

$$p(2) = P(Y = 2) = \frac{\binom{3}{2}\binom{3}{0}}{15} = \frac{3}{15} = \frac{1}{5}.$$

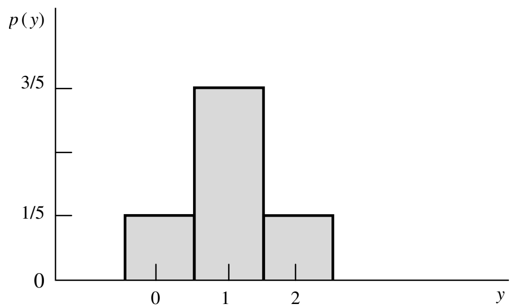
Observe que $(Y = 1)$ es por mucho el resultado más probable. Esto debe parecer razonable porque el número de mujeres es igual al de hombres en el grupo original. ■

La tabla para la distribución de probabilidad de la variable aleatoria Y considerada en el Ejemplo 3.1 se resume en la Tabla 3.1. La misma distribución se da en forma gráfica en la Figura 3.1. Si consideramos el ancho de cada barra de la Figura 3.1 como una unidad, enton-

Tabla 3.1 Distribución de probabilidad para el Ejemplo 3.1

y	$p(y)$
0	1/5
1	3/5
2	1/5

FIGURA **3.1** Histograma de probabilidad para la Tabla 3.1



ces el área de una barra es igual a la probabilidad de que Y tome el valor sobre el cual está centrada la barra. Este concepto de áreas que representan probabilidades se introdujo en la Sección 1.2.

El método más conciso para representar distribuciones de probabilidad discretas es por medio de una fórmula. Para el Ejemplo 3.1, vemos que la fórmula para $p(y)$ se puede escribir como

$$p(y) = \frac{\binom{3}{y}\binom{3}{2-y}}{\binom{6}{2}}, \qquad y = 0,\ 1,\ 2.$$

Observe que las probabilidades asociadas con todos los valores distintos de una variable aleatoria discreta deben sumar 1. En resumen, las siguientes propiedades deben cumplirse para cualquier distribución de probabilidad discreta:

TEOREMA **3.1**

Para cualquier distribución de probabilidad discreta, lo siguiente debe ser verdadero:

1. $0 \leq p(y) \leq 1$ para toda y.
2. $\sum_y p(y) = 1$, donde la sumatoria es para todos los valores de y con probabilidad diferente de cero.

Como ya se mencionó en la Sección 1.5, las distribuciones de probabilidad que dedujimos son *modelos*, no representaciones exactas, por las distribuciones de frecuencia de poblaciones de datos reales que ocurren (o serían generadas) en la naturaleza. Entonces, son modelos para distribuciones reales de datos semejantes a las distribuciones examinadas en el Capítulo 1. Por ejemplo, si fuéramos a seleccionar al azar dos trabajadores de entre los seis descritos en el Ejemplo 3.1, observaríamos un solo valor y. En este caso el valor y observado sería 0, 1 o 2. Si el experimento se repitiera muchas veces, se generarían numerosos valores y. Un histograma de frecuencia relativa para los datos resultantes, construido en la forma descrita en el Capítulo 1, sería muy similar al histograma de probabilidad de la Figura 3.1.

Estos estudios de simulación son muy útiles. Al repetir algunos experimentos una y otra vez podemos generar mediciones de variables aleatorias discretas que tienen distribuciones de frecuencia muy semejantes a las distribuciones de probabilidad deducidas en este capítulo, reforzando la convicción de que nuestros modelos son muy precisos.

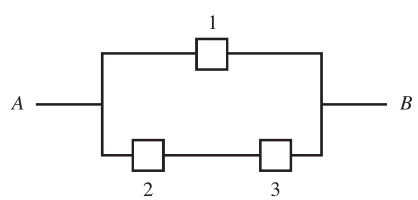
Ejercicios

3.1 Cuando el departamento de salud examinó pozos privados en un condado en busca de dos impurezas que comúnmente se hallan en el agua potable, se encontró que 20% de los pozos no tenían ninguna impureza, 40% tenían la impureza A y 50% tenían la impureza B. (Obviamente, algunos tenían ambas impurezas.) Si un pozo de los existentes en el condado se escoge al azar, encuentre la distribución de probabilidad para Y, el número de impurezas halladas en el pozo.

3.2 Usted y un amigo participan en un juego donde cada uno tira al aire una moneda balanceada. Si las caras superiores de las monedas son cruces en ambos casos, el lector gana \$1; si salen caras en ambos tiros, gana \$2; si las caras de las monedas no son iguales (cara en una y cruz en la otra), el lector pierde \$1 (gana (−\$1)). Obtenga la distribución de probabilidad para sus ganancias, Y, en un solo intento.

3.3 Se sabe que un grupo de cuatro componentes dos de ellos son defectuosos. Una inspectora prueba los componentes uno por uno hasta hallar los dos defectuosos. Una vez que los localiza, suspende la prueba pero el segundo defectuoso es probado para asegurar la precisión. Denote con Y el número de la prueba en la que se halló el segundo componente defectuoso. Encuentre la distribución de probabilidad para Y.

3.4 Considere un sistema de agua que circula por válvulas de A a B. (Vea el diagrama siguiente.) Las válvulas 1, 2 y 3 operan de manera independiente y cada una abre correctamente a una señal con probabilidad .8. Encuentre la distribución de probabilidad para Y, el número de trayectorias abiertas de A a B después que se da la señal. (Observe que Y puede tomar los valores 0, 1 y 2.)



3.5 Un problema en un examen aplicado a niños pequeños les pide relacionar cada una de tres imágenes de animales con la palabra que identifica a ese animal. Si un niño asigna las tres palabras al azar a las tres imágenes, encuentre la distribución de probabilidad para Y, el número de pares correctos.

3.6 Cinco pelotas numeradas 1, 2, 3, 4 y 5 se ponen en una urna. Dos de ellas se seleccionan al azar de entre las cinco y se toma nota de sus números. Encuentre la distribución de probabilidad para lo siguiente:

a El *mayor* de los dos números muestreados.

b La *suma* de los dos números muestreados.

3.7 Cada una de tres bolas se colocan al azar en uno de tres tazones. Encuentre la distribución de probabilidad para Y = el número de tazones vacíos.

3.8 Una sola célula puede morir, con probabilidad .1 o dividirse en dos células, con probabilidad .9, produciendo una nueva generación de células. Cada célula en la nueva generación muere o se divide independientemente en dos células, con las mismas probabilidades que la célula inicial. Encuentre la distribución de probabilidad para el número de células de la siguiente generación.

3.9 Para verificar la exactitud de sus cuentas financieras, las empresas emplean regularmente auditores para verificar asientos de contabilidad. Los empleados de la compañía hacen asientos erróneos 5% de las veces. Suponga que un auditor verifica al azar tres asientos.

a Encuentre la distribución de probabilidad para Y, el número de errores detectados por el auditor.

b Construya un histograma de probabilidad para $p(y)$.

c Encuentre la probabilidad de que el auditor detecte más de un error.

3.10 Una agencia de rentas, que alquila equipo pesado por día, ha encontrado que, en promedio, se renta una costosa máquina sólo un día de cada cinco. Si la renta en un día es independiente de la renta en cualquier otro día, encuentre la distribución de probabilidad de Y, el número de días entre un par de rentas.

3.11 Las personas que entran a un banco de sangre son tales que 1 de cada 3 tienen tipo de sangre O^+ y 1 de cada 15 tienen sangre tipo O^-. Considere tres donantes seleccionados al azar para el banco de sangre. Denote con X el número de donadores con sangre tipo O^+ y denote con Y el número con sangre tipo O^-. Encuentre las distribuciones de probabilidad para X y Y. También encuentre la distribución de probabilidad para $X + Y$, el número de donadores que tienen tipo de sangre O.

3.3 El valor esperado de una variable aleatoria o una función de una variable aleatoria

Hemos observado que la distribución de probabilidad para una variable es un modelo teórico para la distribución empírica de datos asociados con una población real. Si el modelo es una representación precisa de la naturaleza, las distribuciones teóricas y empíricas son equivalentes. En consecuencia, al igual que en el Capítulo 1, tratamos de hallar la media y la varianza para una variable aleatoria y por tanto adquirir medidas numéricas descriptivas, *parámetros*, para la distribución de probabilidad $p(y)$ que son consistentes con las estudiadas en el Capítulo 1.

DEFINICIÓN **3.4**

Sea Y una variable aleatoria discreta con la función de probabilidad $p(y)$. Entonces el *valor esperado* de Y, $E(Y)$, se define como[2]

$$E(Y) = \sum_y yp(y).$$

Si $p(y)$ es una caracterización precisa de la distribución de frecuencia poblacional, entonces $E(Y) = \mu$ es la media poblacional.

La Definición 3.4 es completamente consistente con la definición de la media de un conjunto de mediciones que se dio en la Definición 1.1. Por ejemplo, considere una variable aleatoria discreta Y que puede tomar valores 0, 1 y 2 con distribución de probabilidad $p(y)$ como

2. Para ser precisos, se dice que el valor esperado de una variable aleatoria discreta existe si la suma, como se dio antes, es absolutamente convergente, es decir, si

$$\sum_y |y|\, p(y) < \infty.$$

Esta convergencia absoluta se cumplirá para todos los ejemplos de este texto y no se mencionará cada vez que se defina un valor esperado.

Tabla 3.2 Distribución de probabilidad para *Y*

y	$p(y)$
0	1/4
1	1/2
2	1/4

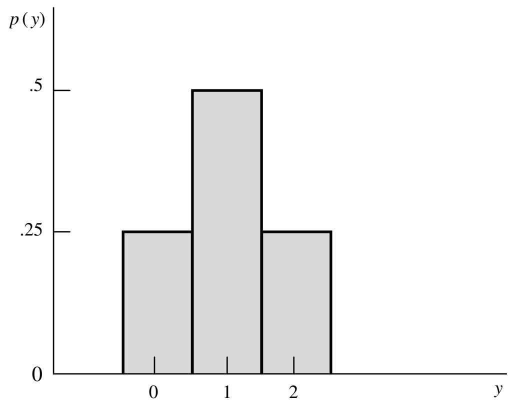


FIGURA **3.2** Distribución de probabilidad para Y

se muestra en la Tabla 3.2 y el histograma de probabilidad de la Figura 3.2. Una inspección visual revelará que la media de la distribución está ubicada en $y = 1$.

Para demostrar que $E(Y) = \sum_y yp(y)$ es la media de la distribución de probabilidad $p(y)$, suponga que el experimento se realizó 4 millones de veces, dando 4 millones de valores observados para Y. Si observamos $p(y)$ de la Figura 3.2, esperaríamos que *aproximadamente* 1 millón de los 4 millones de repeticiones se manifieste en el resultado $Y = 0$, 2 millones en $Y = 1$ y 1 millón en $Y = 2$. Para hallar el valor medio de Y, promediamos estos 4 millones de mediciones y obtenemos

$$\begin{aligned}\mu \approx \frac{\sum_{i=1}^{n} y_i}{n} &= \frac{(1{,}000{,}000)(0) + (2{,}000{,}000)(1) + (1{,}000{,}000)(2)}{4{,}000{,}000} \\ &= (0)(1/4) + (1)(1/2) + (2)(1/4) \\ &= \sum_{y=0}^{2} yp(y) = 1.\end{aligned}$$

Entonces, $E(Y)$ es un promedio, y la Definición 3.4 es consistente con la definición de una media dada en la Definición 1.1. Del mismo modo, frecuentemente estamos interesados en la media o el valor esperado de una función de una variable aleatoria Y. Por ejemplo, las moléculas en el espacio se mueven a velocidades que varían, donde Y, la velocidad de una molécula determinada, es una variable aleatoria. La energía transmitida al impactarse con un cuerpo en movimiento es proporcional al cuadrado de la velocidad. En consecuencia, para hallar la cantidad media de energía transmitida por una molécula en el impacto, debemos hallar el valor medio de Y^2. Lo que es más importante, observamos en la Definición 1.2 que la varianza de un conjunto de mediciones es la media del cuadrado de las diferencias entre cada valor del conjunto de mediciones y su media o el valor medio de $(Y - \mu)^2$.

TEOREMA 3.2

Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sea $g(Y)$ una función de valor real de Y. Entonces, el valor esperado de $g(Y)$ está dado por

$$E[g(Y)] = \sum_{\text{toda } y} g(y)p(y).$$

Demostración

Demostramos el resultado en el caso que la variable aleatoria Y toma el número finito de valores $y_1, y_2, \ldots, y_n$. Como la función $g(y)$ puede no ser biunívoca, suponga que $g(Y)$ toma valores $g_1, g_2, \ldots, g_m$ (donde $m \leq n$). Se deduce que $g(Y)$ es una variable aleatoria tal que para $i = 1, 2, \ldots, m$,

$$P[g(Y) = g_i] = \sum_{\substack{\text{toda } y_j \text{ tal que} \\ g(y_j) = g_i}} p(y_j) = p^*(g_i).$$

Entonces, por la Definición 3.4,

$$\begin{aligned} E[g(Y)] &= \sum_{i=1}^{m} g_i p^*(g_i) \\ &= \sum_{i=1}^{m} g_i \left\{ \sum_{\substack{\text{toda } y_j \text{ tal que} \\ g(y_j) = g_i}} p(y_j) \right\} \\ &= \sum_{i=1}^{m} \sum_{\substack{\text{toda } y_j \text{ tal que} \\ g(y_j) = g_i}} g_i p(y_j) \\ &= \sum_{j=1}^{n} g(y_j)p(y_j). \end{aligned}$$

Regresemos ahora a nuestro objetivo inmediato, hallar medidas numéricas descriptivas (o *parámetros*) para caracterizar $p(y)$. Como ya dijimos, $E(Y)$ proporciona la media de la población con distribución dada por $p(y)$. A continuación buscamos la varianza y la desviación estándar de esta población. El lector recordará del Capítulo 1 que la varianza de un conjunto de mediciones es el promedio del cuadrado de las diferencias entre los valores de un conjunto de mediciones y su media. Entonces, deseamos hallar el valor medio de la función $g(Y) = (Y - \mu)^2$.

DEFINICIÓN 3.5

Si Y es una variable aleatoria con media $E(Y) = \mu$, la varianza de una variable aleatoria Y se define como el valor esperado de $(Y - \mu)^2$. Esto es,

$$V(Y) = E[(Y - \mu)^2].$$

La *desviación estándar* de Y es la raíz cuadrada positiva de $V(Y)$.

Si $p(y)$ es una caracterización precisa de la distribución de frecuencia poblacional (y para simplificar la notación, supondremos que esto es cierto), entonces $E(Y) = \mu$, $V(Y) = \sigma^2$, la varianza poblacional, y σ es la desviación estándar poblacional.

EJEMPLO **3.2** La distribución de probabilidad para una variable aleatoria Y está dada en la Tabla 3.3. Encuentre la media, la varianza y la desviación estándar de Y.

Tabla 3.3 Distribución de probabilidad para *Y*

y	$p(y)$
0	1/8
1	1/4
2	3/8
3	1/4

Solución Por las Definiciones 3.4 y 3.5,

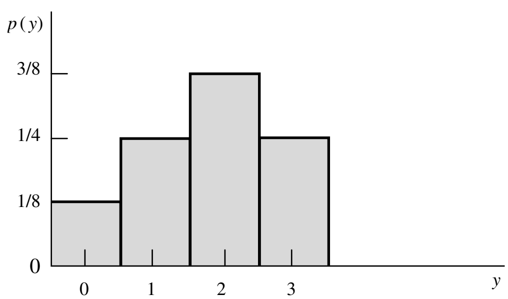
$$\mu = E(Y) = \sum_{y=0}^{3} yp(y) = (0)(1/8) + (1)(1/4) + (2)(3/8) + (3)(1/4) = 1.75,$$

$$\begin{aligned}\sigma^2 &= E[(Y-\mu)^2] = \sum_{y=0}^{3}(y-\mu)^2 p(y)\\ &= (0-1.75)^2(1/8) + (1-1.75)^2(1/4) + (2-1.75)^2(3/8) + (3-1.75)^2(1/4)\\ &= .9375,\end{aligned}$$

$$\sigma = +\sqrt{\sigma^2} = \sqrt{.9375} = .97.$$

El histograma de probabilidad se muestra en la Figura 3.3. Localice μ en el eje de medición y observe que ubica el "centro" de la distribución de probabilidad no simétrica de Y. También observe que el intervalo $(\mu \pm \sigma)$ contiene los puntos discretos $Y = 1$ y $Y = 2$, que constituyen $5/8$ de la probabilidad. Así, la regla empírica (Capítulo 1) proporciona una aproximación razonable a la probabilidad de una medición que cae en este intervalo. (Recuerde que las probabilidades se concentran en los puntos $Y = 0, 1, 2$ y 3 porque Y no puede tomar valores intermedios.)

FIGURA **3.3** Histograma de probabilidad para el Ejemplo 3.2



■

Será útil adquirir unas pocas herramientas y definiciones adicionales antes de tratar de hallar los valores y varianzas esperados de variables aleatorias discretas más complicadas, por ejemplo la binomial o de Poisson. Es así que presentamos tres útiles teoremas de expectativa que se deducen directamente de la teoría de sumatoria. (Otras técnicas útiles se presentan en las Secciones 3.4 y 3.9.) Para cada teorema suponemos que Y es una variable aleatoria discreta con función de probabilidad $p(y)$.

El primer teorema expresa el resultado más bien obvio de que la media o valor esperado de una cantidad no aleatoria c es igual a c.

TEOREMA 3.3 Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sea c una constante. Entonces $E(c) = c$.

Demostración Considere la función $g(Y) \equiv c$. Por el Teorema 3.2,

$$E(c) = \sum_y cp(y) = c \sum_y p(y).$$

Pero $\sum_y p(y) = 1$ (Teorema 3.1) y, por tanto, $E(c) = c(1) = c$.

El segundo teorema expresa que el valor esperado del producto de una constante c por una función de una variable aleatoria es igual a la constante multiplicada por el valor esperado de la función de la variable.

TEOREMA 3.4 Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$, $g(Y)$ una función de Y y c una constante. Entonces

$$E[cg(Y)] = cE[g(Y)].$$

Demostración Por el Teorema 3.2,

$$E[cg(Y)] = \sum_y cg(y)p(y) = c \sum_y g(y)p(y) = cE[g(Y)].$$

El tercer teorema expresa que la media o valor esperado de una suma de funciones de una variable aleatoria Y es igual a la suma de sus respectivos valores esperados.

TEOREMA 3.5 Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y sean $g_1(Y), g_2(Y), \dots,$ $g_k(Y)$ k funciones de Y. Entonces

$$E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)].$$

Demostración Demostraremos la prueba sólo para el caso $k = 2$, pero pasos análogos se cumplirán para cualquier k finita. Por el Teorema 3.2,

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)]. \end{aligned}$$

Los Teoremas 3.3, 3.4 y 3.5 se pueden usar de inmediato para desarrollar un teorema útil para hallar la varianza de una variable aleatoria discreta.

TEOREMA **3.6** Sea Y una variable aleatoria discreta con función de probabilidad $p(y)$ y media $E(y) = \mu$; entonces

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2.$$

Demostración

$$\begin{aligned}\sigma^2 &= E[(Y - \mu)^2] = E(Y^2 - 2\mu Y + \mu^2)\\ &= E(Y^2) - E(2\mu Y) + E(\mu^2) \qquad \text{(por el Teorema 3.5).}\end{aligned}$$

Si se observa que μ es una constante y se aplican los Teoremas 3.4 y 3.3 a los términos segundo y tercero, respectivamente, tenemos

$$\sigma^2 = E(Y^2) - 2\mu E(Y) + \mu^2.$$

Pero $\mu = E(Y)$ y, por tanto,

$$\sigma^2 = E(Y^2) - 2\mu^2 + \mu^2 = E(Y^2) - \mu^2.$$

En ocasiones el Teorema 3.6 reduce considerablemente el trabajo de hallar la varianza de una variable aleatoria discreta. Demostraremos la utilidad de este resultado al calcular de nuevo la varianza de la variable aleatoria considerada en el Ejemplo 3.2.

EJEMPLO **3.3** Use el Teorema 3.6 para hallar la varianza de la variable aleatoria Y en el Ejemplo 3.2.

Solución La media $\mu = 1.75$ se encontró en el Ejemplo 3.2. Como

$$E(Y^2) = \sum_y y^2 p(y) = (0)^2(1/8) + (1)^2(1/4) + (2)^2(3/8) + (3)^2(1/4) = 4,$$

el Teorema 3.6 dice que

$$\sigma^2 = E(Y^2) - \mu^2 = 4 - (1.75)^2 = .9375.$$

■

EJEMPLO **3.4** El gerente de una planta industrial está planeando comprar una nueva máquina ya sea del tipo A o del B. Si t denota el número de horas de operación diaria, el número Y_1 de reparaciones diarias requeridas para mantener una máquina de tipo A es una variable aleatoria con media y varianza iguales a $.10t$. El número Y_2 de reparaciones diarias para una máquina de tipo B es una variable aleatoria con media y varianza iguales a $.12t$. El costo diario de operación A es $C_A(t) = 10t + 30Y_1^2$; para B es $C_B(t) = 8t + 30Y_2^2$. Suponga que las reparaciones toman un tiempo insignificante y que cada noche las máquinas se afinan para que puedan operar esencialmente como máquinas nuevas al empezar el día siguiente. ¿Cuál máquina minimiza el costo diario esperado si un día de trabajo consta de (a) 10 horas y (b) 20 horas?

Solución El costo diario esperado para A es

$$E[C_A(t)] = E\left[10t + 30Y_1^2\right] = 10t + 30E\left(Y_1^2\right)$$
$$= 10t + 30\{V(Y_1) + [E(Y_1)]^2\} = 10t + 30[.10t + (.10t)^2]$$
$$= 13t + .3t^2.$$

En este cálculo, usamos los valores conocidos para $V(Y_1)$ y $E(Y_1)$ y el hecho de que $V(Y_1) = E(Y_1^2) - [E(Y_1)]^2$ para obtener que $E(Y_1^2) = V(Y_1) + [E(Y_1)]^2 = .10t + (.10t)^2$. De manera similar,

$$E[C_B(t)] = E\left[8t + 30Y_2^2\right] = 8t + 30E\left(Y_2^2\right)$$
$$= 8t + 30\{V(Y_2) + [E(Y_2)]^2\} = 8t + 30[.12t + (.12t)^2]$$
$$= 11.6t + .432t^2.$$

Por tanto, para la situación (a) donde $t = 10$,

$$E[C_A(10)] = 160 \qquad \text{y} \qquad E[C_B(10)] = 159.2,$$

resulta en la selección de la máquina B.

Para la situación (b), $t = 20$ y

$$E[C_A(20)] = 380 \qquad \text{y} \qquad E[C_B(20)] = 404.8,$$

resultando en la selección de la máquina A.

En conclusión, las máquinas de tipo B son más económicas para periodos cortos por su menor costo de operación por hora. No obstante, para periodos largos, las máquinas de tipo A son más económicas porque tienden a ser reparadas con menos frecuencia. ■

El propósito de esta sección fue introducir el concepto de un valor esperado y desarrollar algunos teoremas útiles para hallar medias y varianzas de variables aleatorias o funciones de variables aleatorias. En las secciones siguientes presentamos algunos tipos específicos de variables aleatorias discretas y damos fórmulas para sus distribuciones de probabilidad y sus medias y varianzas. Como verá, en realidad deducir algunos de estos valores esperados requiere de experiencia en la suma de series algebraicas y conocimiento de unos cuantos trucos. Ilustraremos algunos de ellos en varias de las deducciones de las secciones siguientes.

Ejercicios

3.12 Sea Y una variable aleatoria con $p(y)$ dada en la tabla siguiente. Encuentre $E(Y)$, $E(1/Y)$, $E(Y^2 - 1)$ y $V(Y)$.

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

3.13 Consulte el juego de tiro al aire de una moneda del Ejercicio 3.2. Calcule la media y la varianza de Y, que son sus ganancias en un solo intento. Observe que $E(Y) > 0$. ¿Cuánto debe pagar usted por participar en este juego si sus *ganancias netas*, es decir, la diferencia entre pago y costo de jugar han de tener una media de 0?

3.14 La vida máxima de la patente para un nuevo medicamento es 17 años. Si restamos el tiempo requerido por la FDA para someter a pruebas y aprobar el medicamento, se obtiene la vida real de la patente para el medicamento, es decir, el tiempo que la compañía tiene para recuperar los costos de investigación y desarrollo y para obtener una utilidad. La distribución de los tiempos de vida reales de las patentes para nuevos medicamentos se da a continuación:

Años, y	3	4	5	6	7	8	9	10	11	12	13
$p(y)$	.03	.05	.07	.10	.14	.20	.18	.12	.07	.03	.01

a Encuentre la vida media de la patente para un nuevo medicamento.

b Encuentre la desviación estándar de $Y =$ tiempo de vida de un nuevo medicamento seleccionado al azar.

c ¿Cuál es la probabilidad de que el valor de Y caiga en el intervalo $\mu \pm 2\sigma$?

3.15 ¿Quién es el rey de los programas de TV por la noche? Una encuesta en Internet estima que, cuando se les da a elegir entre David Letterman y Jay Leno, 52% de la población prefiere ver a Jay Leno. Al azar se seleccionan tres personas que ven TV hasta tarde y se les pregunta cuál de los dos presentadores de programas prefieren.

a Encuentre la distribución de probabilidad para Y, el número de personas de la muestra que prefieren a Leno.

b Construya un histograma de probabilidad para $p(y)$.

c ¿Cuál es la probabilidad de que exactamente una de esas tres personas prefiera a Leno?

d ¿Cuáles son la media y la desviación estándar para Y?

e ¿Cuál es la probabilidad de que el número de personas a favor de Leno caiga a no más de 2 desviaciones estándar de la media?

3.16 Al secretario del Ejercicio 2.121 se le dieron n contraseñas de computadora e intenta usarlas al azar. Exactamente una contraseña permitirá el acceso a un archivo de computadora. Encuentre la media y la varianza de Y, que es el número de intentos necesarios para abrir el archivo, si se eliminan las contraseñas no exitosas (como en el Ejercicio 2.121).

3.17 Consulte el Ejercicio 3.7. Encuentre la media y la desviación estándar para $Y =$ número de tazones vacíos. ¿Cuál es la probabilidad de que el valor de Y caiga a no más de 2 desviaciones estándar de la media?

3.18 Consulte el Ejercicio 3.8. ¿Cuál es el número medio de células de la segunda generación?

3.19 Una compañía de seguros expide una póliza de un año por \$1000 dólares contra el suceso A que históricamente le ocurre a 2 de cada 100 propietarios de la póliza. Las tarifas administrativas son de \$15 por póliza y no son parte de la "utilidad" de la compañía. ¿Cuánto debe cobrar la compañía por la póliza si requiere que la utilidad esperada por póliza sea de \$50? [*Sugerencia:* si C es la prima por la póliza, la "utilidad" de la compañía es $C - 15$ si A no ocurre y $C - 15 - 1000$ si A ocurre.]

3.20 Una compañía manufacturera envía su producto en camiones con remolques de dos tamaños diferentes. Cada embarque se hace en un remolque con dimensiones de 8 pies $\times$ 10 pies $\times$ 30 pies o de 8 pies $\times$ 10 pies $\times$ 40 pies. Si 30% de sus envíos se hacen usando remolques de 30 pies y 70% en remolques de 40 pies, encuentre el volumen medio enviado en cada remolque. (Suponga que los remolques siempre están llenos.)

3.21 El número N de casas residenciales a las que una compañía de bomberos da servicio depende de la distancia r (en manzanas) que una motobomba puede alcanzar en un tiempo especificado (fijo). Si suponemos que N es proporcional al área de un círculo de R manzanas desde la estación de bomberos, entonces $N = C\pi R^2$, donde C es una constante, $\pi = 3.1416...$, y R, la variable aleatoria, es el número de manzanas que una motobomba se puede trasladar en el tiempo especificado. Para una compañía particular de bomberos, $C = 8$, la distribución de probabilidad para R es como se muestra en la tabla siguiente y $p(r) = 0$ para $r \leq 20$ y $r \geq 27$.

r	21	22	23	24	25	26
$p(r)$	.05	.20	.30	.25	.15	.05

Encuentre el valor esperado de N, el número de casas a las que el departamento de bomberos puede atender.

3.22 Un dado balanceado se tira una vez. Sea Y el número en su cara superior. Encuentre el valor esperado y la varianza de Y.

3.23 En un juego de azar una persona saca una sola carta de una baraja ordinaria de 52 cartas. A una persona le pagan \$15 por sacar una "sota" o una reina y \$5 por sacar un rey o un as. Alguien que saque cualquier otra carta paga \$4. Si una persona participa en este juego, ¿cuál es la ganancia esperada?

3.24 Aproximadamente 10% de las botellas de vidrio que salen de una línea de producción presentan defectos serios en el vidrio. Si dos botellas se seleccionan al azar, encuentre la media y la varianza del número de botellas que presentan defectos serios.

3.25 Dos contratos de construcción se van a asignar al azar a una o más de tres empresas: I, II y III. Cualquier empresa puede recibir ambos contratos. Si cada contrato dará una utilidad de \$90,000 para la empresa, encuentre la utilidad esperada para la empresa I. Si las empresas I y II son propiedad de la misma persona, ¿cuál es la utilidad esperada total del propietario?

***3.26** Un vendedor de equipo pesado puede comunicarse con uno o dos clientes por día con probabilidades $1/3$ y $2/3$, respectivamente. Cada contacto resultará ya sea en que no haya venta o en una venta de \$50,000, con las probabilidades .9 y .1, respectivamente. Obtenga la distribución de probabilidad para ventas al día. Encuentre la media y la desviación estándar de las ventas al día.[3]

3.27 Un cliente potencial para una póliza de seguro contra incendio por \$85,000 es propietario de una casa en una zona que, de acuerdo con la experiencia, puede sostener una pérdida total en un año determinado con probabilidad de .001 y un 50% de pérdida con probabilidad .01. Si se ignoran todas las otras pérdidas parciales, ¿qué prima debe cobrar la compañía de seguros por una póliza anual para que no haya pérdida ni ganancia en todas las pólizas de \$85,000 en esta zona?

3.28 Consulte el Ejercicio 3.3. Si el costo de probar un componente es \$2 y el costo de reparar uno defectuoso es \$4, encuentre el costo total esperado por probar y reparar el lote.

***3.29** Si Y es una variable aleatoria discreta que asigna probabilidades positivas a sólo los enteros positivos, demuestre que

$$E(Y) = \sum_{i=1}^{\infty} P(Y \geq k).$$

3.30 Suponga que Y es una variable aleatoria discreta con media μ y varianza σ^2 y sea $X = Y + 1$.

a ¿Se espera que la media de X sea mayor, menor o igual que $\mu = E(Y)$? ¿Por qué?

b Use los Teoremas 3.3 y 3.5 para expresar $E(X) = E(Y + 1)$ en términos de $\mu = E(Y)$. ¿Este resultado está de acuerdo con la respuesta al inciso (a)?

c Recordando que la varianza es una medida de dispersión, ¿espera usted que la varianza de X sea mayor, menor o igual que $\sigma^2 = V(Y)$? ¿Por qué?

3. Los ejercicios precedidos por un asterisco son opcionales.

d Use la Definición 3.5 y el resultado del inciso b para demostrar que

$$V(X) = E\{[(X - E(X)]^2\} = E[(Y - \mu)^2] = \sigma^2;$$

esto es, $X = Y + 1$ y Y tienen varianzas *iguales*.

3.31 Suponga que Y es una variable aleatoria discreta con media μ y varianza σ^2 y sea $W = 2Y$.

a ¿Espera usted que la media de W sea mayor, menor o igual que $\mu = E(Y)$? ¿Por qué?

b Use el Teorema 3.4 para expresar $E(W)= E(2Y)$ en términos de $\mu = E(Y)$. ¿Este resultado está de acuerdo con la respuesta al inciso a?

c Recordando que la varianza es una medida de dispersión, ¿espera usted que la varianza de W sea mayor, menor o igual que $\sigma^2 = V(Y)$? ¿Por qué?

d Use la Definición 3.5 y el resultado del inciso b para demostrar que

$$V(W) = E\{[W - E(W)]^2\} = E[4(Y - \mu)^2] = 4\sigma^2;$$

esto es, $W = 2Y$ tiene una varianza que cuadriplica la de Y.

3.32 Suponga que Y es una variable aleatoria discreta con media μ y varianza σ^2 y sea $U = Y/10$.

a ¿Espera usted que la media de U sea mayor, menor o igual a $\mu = E(Y)$? ¿Por qué?

b Use el Teorema 3.4 para expresar $E(U)= E(Y/10)$ en términos que $\mu = E(Y)$. ¿Este resultado está de acuerdo con la respuesta al inciso a?

c Recordando que la varianza es una medida de dispersión, ¿espera usted que la varianza de U sea mayor, menor o igual que $\sigma^2 = V(Y)$? ¿Por qué?

d Use la Definición 3.5 y el resultado del inciso (b) para demostrar que

$$V(U) = E\{[U - E(U)]^2\} = E[.01(Y - \mu)^2] = .01\sigma^2;$$

esto es, $U = Y/10$ tiene varianza .01 veces la de Y.

3.33 Sea Y una variable aleatoria discreta con media μ y varianza σ^2. Si a y b son constantes, use los Teoremas 3.3 y 3.6 para demostrar que

a $E(aY + b) = aE(Y) + b = a\mu + b$,

b $V(aY + b) = a^2V(Y) = a^2\sigma^2$.

3.34 El gerente del almacén de una fábrica ha construido la siguiente distribución de probabilidad para la demanda diaria (número de veces que se usa) de una herramienta en particular.

y	0	1	2
$p(y)$	.1	.5	.4

Cuesta a la fábrica \$10 cada vez que la herramienta se usa. Encuentre la media y la varianza del costo diario por usar la herramienta.

3.4 La distribución de probabilidad binomial

Algunos experimentos consisten en la observación de una secuencia de intentos idénticos e independientes, cada uno de los cuales puede resultar en una de dos salidas. Cada artículo que sale de la línea de producción de manufacturas es defectuoso o no defectuoso. Cada disparo en una secuencia de tiros a un blanco puede resultar en un acierto o en no acierto y cada una de las n personas entrevistadas antes de una elección local está a favor del candidato Jones o no lo está. En esta sección estamos interesados en experimentos, conocidos como *experimentos binomiales,* que presentan las siguientes características.

DEFINICIÓN **3.6**

Un *experimento binomial* presenta las siguientes propiedades:

1. Consiste en un número fijo, n, de pruebas idénticas.
2. Cada prueba resulta en uno de dos resultados: éxito, S, o fracaso, F.
3. La probabilidad de éxito en una sola prueba es igual a algún valor p y es el mismo de una prueba a la otra. La probabilidad de fracaso es igual a $q = (1 - p)$.
4. Las pruebas son independientes.
5. La variable aleatoria de interés es Y, el número de éxitos observado durante las n pruebas.

Determinar si un experimento particular es binomial requiere examinar el experimento en busca de cada una de las características recién enlistadas. Observe que la variable aleatoria de interés es el número de éxitos observado en las n pruebas. Es importante ver que un éxito no es necesariamente "bueno" en el sentido diario de la palabra. En nuestras exposiciones, éxito es simplemente un nombre para uno de los dos posibles resultados en una sola prueba de un experimento.

EJEMPLO **3.5** Un sistema de detección de alarma temprana para aviones consta de cuatro unidades de radar idénticas que operan de manera independiente entre sí. Suponga que cada una tiene una probabilidad de .95 de detectar un avión intruso. Cuando un avión intruso entra en escena, la variable aleatoria de interés es Y, el número de unidades de radar que *no detecta* el avión. ¿Es éste un experimento binomial?

Solución Para decidir si éste es un experimento binomial, debemos determinar si cada uno de los cinco requisitos de la Definición 3.6 se satisface. Observe que la variable aleatoria de interés Y, es el número de unidades de radar que *no* detecta un avión. La variable aleatoria de interés en un experimento binomial es siempre el número de éxitos; en consecuencia, este experimento puede ser binomial sólo si llamamos éxito al evento de *no detectar*. Ahora examinemos el experimento en cuanto a las cinco características del experimento binomial.

1. El experimento comprende cuatro pruebas idénticas; cada una de ellas consiste en determinar si una unidad particular de radar detecta (o no) el avión.
2. Cada prueba arroja uno de dos resultados. Como la variable aleatoria de interés es el número de éxitos, S denota que el avión no fue detectado y F denota que fue detectado.
3. Como todas las unidades de radar detectan el avión con igual probabilidad, la probabilidad de una S en cada prueba es la misma y $p = P(S) = P$ (no detectar) $= .05$.

4. Las pruebas son independientes porque las unidades operan de manera independiente.
5. La variable aleatoria de interés es Y, el número de éxitos en cuatro pruebas.

Entonces, el experimento es binomial con $n = 4$, $p = .05$ y $q = 1 - .05 = .95$. ■

EJEMPLO 3.6 Suponga que 40% de una población grande de votantes registrados está a favor del candidato Jones. Una muestra aleatoria de $n = 10$ votantes será seleccionada y se ha de observar Y, el número de quienes están a favor de Jones. ¿El experimento satisface los requisitos de un experimento binomial?

Solución Si cada una de las diez personas se selecciona al azar de entre la población, entonces tenemos diez pruebas casi idénticas con cada prueba resultando en una persona que está a favor de Jones (S) o no está a favor de Jones (F). La variable aleatoria de interés es entonces el número de éxitos en las diez pruebas. Para la primera persona seleccionada, la probabilidad de estar a favor de Jones (S) es .4. Pero, ¿qué se puede decir acerca de la probabilidad *incondicional* de que la segunda persona esté a favor de Jones? En el Ejercicio 3.35 se demostrará que la probabilidad *incondicional* de que la segunda persona esté a favor de Jones también es .4. Entonces, la probabilidad de un éxito S sigue siendo la misma de una prueba a otra. No obstante, la probabilidad *condicional* de un éxito en pruebas posteriores depende del número de éxitos en las pruebas previas. Si la población de electores es grande, eliminar una persona no cambia en forma importante la fracción de electores que están a favor de Jones y la probabilidad *condicional* de que la segunda persona esté a favor de Jones será muy cercana a .4. En general, si la población es grande y el tamaño muestral es relativamente pequeño, la probabilidad *condicional* de éxito en una prueba posterior dado el número de éxitos en las pruebas previas seguirá siendo más o menos igual cualesquiera que sean los resultados en pruebas previas. Por tanto, las pruebas serán aproximadamente independientes y entonces los problemas de muestreo de este tipo son aproximadamente binomiales. ■

Si el tamaño muestral del Ejemplo 3.6 fue grande con respecto al tamaño poblacional (por ejemplo 10% de la población), la probabilidad *condicional* de seleccionar un partidario de Jones en una elección posterior será alterada de manera importante por las preferencias de las personas seleccionadas antes en el experimento y éste no sería binomial. La distribución de probabilidad hipergeométrica, tema de la Sección 3.7, es el modelo apropiado de probabilidad a usarse cuando el tamaño muestral es grande con respecto al tamaño poblacional.

Usted puede refinar su capacidad de identificar experimentos binomiales si reexamina los ejercicios del final del Capítulo 2. Varios de los experimentos en esos ejercicios son experimentos binomiales o aproximadamente binomiales.

La distribución $p(y)$ de probabilidad binomial se puede deducir al aplicar el método de punto muestral para hallar la probabilidad de que el experimento produzca y éxitos. Cada punto muestral del espacio muestral puede estar caracterizado por un múltiplo n (rearreglo) que contiene las letras S y F, correspondientes a éxito y fracaso. Un punto muestral típico aparecería entonces como

$$\underbrace{SSFSFFFSFS\ldots FS}_{n \text{ posiciones}},$$

donde la letra en la i-ésima posición (de izquierda a derecha) indica el resultado de la $i$ésima prueba.

Consideremos ahora un punto muestral particular correspondiente a y éxitos y por tanto contenido en el evento numérico $Y = y$. Este punto muestral,

$$\underbrace{SSSSS\ldots SSS}_{y}\underbrace{FFF\ldots FF}_{n-y},$$

representa la intersección de n *eventos independientes* (los resultados de las n pruebas), en las que hubo y éxitos seguidos por $(n - y)$ fracasos. Como las pruebas fueron independientes y la probabilidad de S, p, sigue igual de una prueba a otra, la probabilidad de este punto muestral es

$$\underbrace{ppppp\cdots ppp}_{\text{términos } y}\ \underbrace{qqq\cdots qq}_{\text{términos } n-\ y} = p^y q^{n-y}.$$

Cada uno de los puntos muestrales del evento $Y = y$ se puede representar como un rearreglo que contenga un número y de éxitos S y $(n - y)$ fracasos. Cualquier punto muestral también tiene probabilidad $p^y q^{n-y}$. Como el número de arreglos que contienen la cantidad y de éxitos S y de $(n - y)$ fracasos F es (del Teorema 2.3)

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

se deduce que el evento $(Y = y)$ está formado por $\binom{n}{y}$ puntos muestrales, cada uno con probabilidad $p^y q^{n-y}$, y que $p(y) = \binom{n}{y} p^y q^{n-y}$, $y = 0, 1, 2, \ldots, n$. El resultado que hemos deducido es la fórmula para la *distribución de probabilidad binomial.*

DEFINICIÓN **3.7**

Se dice que una variable aleatoria Y tiene una *distribución binomial* basada en n pruebas con probabilidad p de éxito si y sólo si

$$p(y) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n \text{ y } 0 \le p \le 1.$$

La Figura 3.4 representa $p(y)$ gráficamente como histogramas de probabilidad, el primero para $n = 10, p = .1$; el segundo para $n = 10, p = .5$; y el tercero para $n = 20, p = .5$. Antes de continuar, reconsideremos la representación para los puntos muestrales de este experimento. Hemos visto que un punto muestral puede estar representado por una secuencia de n letras, cada una de las cuales es S o F. Si el punto muestral contiene exactamente una S, la probabilidad asociada con ese punto muestral es pq^{n-1}. Si otro punto muestral contiene 2 letras S y $(n - 2)$ letras F la probabilidad de este punto muestral es p^2q^{n-2}. Observe que los puntos muestrales para un experimento binomial no son equiparables a menos que $p = .5$.

El término *experimento binomial* se deriva del hecho de que cada prueba arroja *uno de dos* posibles resultados y de que las probabilidades $p(y)$, $y = 0, 1, 2, \ldots, n$, son términos de la

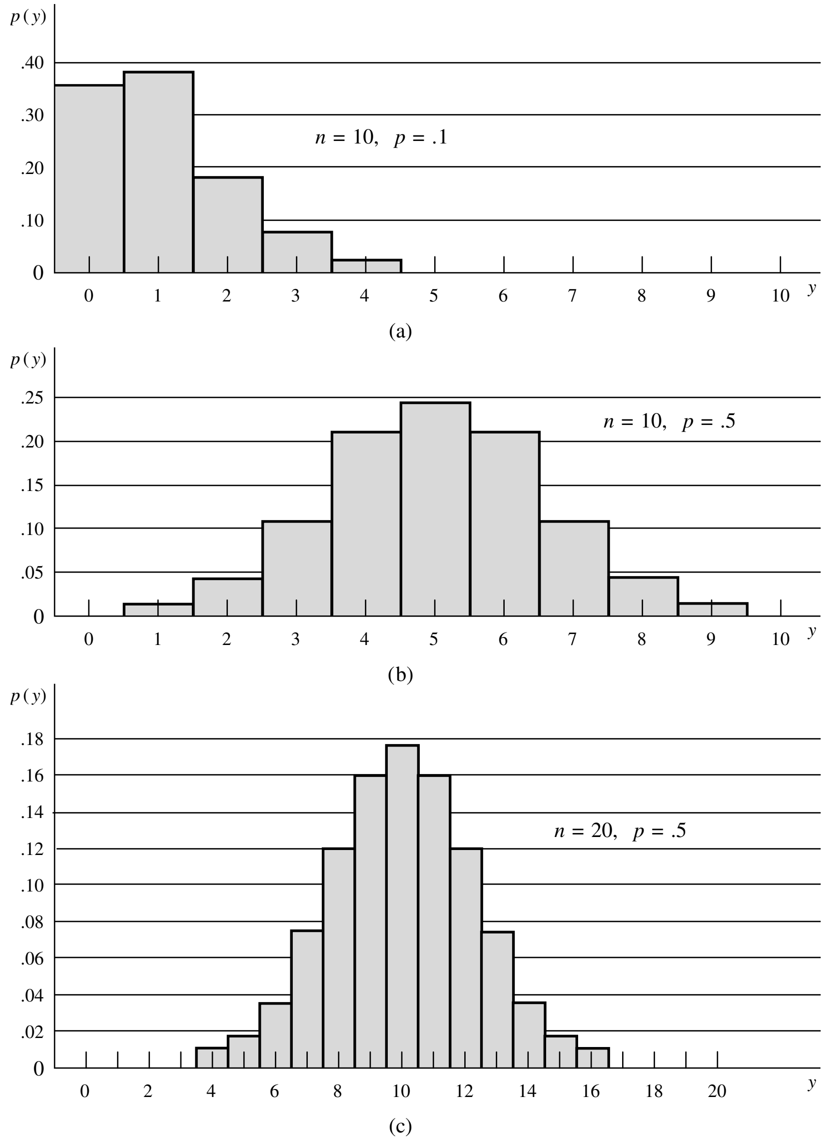


FIGURA 3.4 Histogramas de probabilidad binomial

expansión binomial

$$(q+p)^n = \binom{n}{0}q^n + \binom{n}{1}p^1q^{n-1} + \binom{n}{2}p^2q^{n-2} + \cdots + \binom{n}{n}p^n.$$

Se observa que $\binom{n}{0}q^n = p(0)$, $\binom{n}{1}p^1q^{n-1} = p(1)$ y, en general, $p(y) = \binom{n}{y}p^yq^{n-y}$. También se cumple que $p(y)$ satisface las propiedades necesarias para una función de probabilidad porque $p(y)$ es positiva para $y = 0, 1, \ldots, n$ y [porque $(q+p) = 1$]

$$\sum_y p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y q^{n-y} = (q+p)^n = 1^n = 1.$$

La distribución de probabilidad binomial tiene muchas aplicaciones porque el experimento binomial se presenta al muestrear defectos en control de calidad industrial, en el muestreo de preferencia de los consumidores o en poblaciones de votantes y en muchas otras situaciones físicas. Ilustraremos con algunos ejemplos. Otros modelos prácticos aparecerán en los ejercicios al final de esta sección y también al final del capítulo.

EJEMPLO **3.7** Suponga que un lote de 5000 fusibles eléctricos contiene 5% de piezas defectuosas. Si se prueba una muestra de 5 fusibles, encuentre la probabilidad de hallar al menos uno defectuoso.

Solución Es razonable suponer que Y, el número observado de defectuosos, tiene una distribución binomial aproximada porque el lote es grande. Retirar unos cuantos fusibles no cambia lo suficiente la composición de los restantes como para preocuparnos. Entonces,

$$\begin{aligned} P(\text{al menos uno defectuoso}) &= 1 - p(0) = 1 - \binom{5}{0} p^0 q^5 \\ &= 1 - (.95)^5 = 1 - .774 = .226. \end{aligned}$$

Observe que hay una probabilidad más bien grande de ver al menos uno defectuoso, aun cuando la muestra sea muy pequeña. ■

EJEMPLO **3.8** La experiencia ha demostrado que 30% de todas las personas afectadas por cierta enfermedad se recuperan. Una empresa fabricante de medicamentos ha inventado una nueva medicina. Diez personas con la enfermedad se seleccionaron al azar y recibieron la medicina; nueve se recuperaron al poco tiempo. Suponga que la medicina no es eficaz en absoluto. ¿Cuál es la probabilidad de que se recuperen al menos nueve de entre diez que recibieron la medicina?

Solución Denote con Y el número de personas que se recuperan. Si la medicina no funciona, la probabilidad de que una sola persona enferma se recupere es $p = .3$. Entonces el número de pruebas es $n = 10$ y la probabilidad de que *exactamente* nueve se recuperen es

$$P(Y = 9) = p(9) = \binom{10}{9}(.3)^9(.7) = .000138.$$

Del mismo modo, la probabilidad de que exactamente diez se recuperen es

$$P(Y = 10) = p(10) = \binom{10}{10}(.3)^{10}(.7)^0 = .000006,$$

y

$$P(Y \geq 9) = p(9) + p(10) = .000138 + .000006 = .000144.$$

Si la medicina no es eficaz, la probabilidad de observar al menos nueve que se recuperen es sumamente pequeña. Si administramos la medicina a diez personas y observamos que al menos nueve se recuperan, entonces (1) la medicina no sirve y hemos observado un evento raro o bien (2) la medicina es en verdad útil para curar la enfermedad. Nos apegamos a la conclusión 2. ■

Una tabulación de probabilidades binomiales de la forma $\sum_{y=0}^{a} p(y)$, presentada en la Tabla 1, Apéndice 3, reducirá en gran medida los cálculos para algunos de los ejercicios. Las obras de consulta que aparecen al final del capítulo citan varias tabulaciones más amplias de probabilidades binomiales. Debido a limitaciones prácticas de espacio, las tablas impresas por lo general se aplican sólo a valores seleccionados de n y p. Las probabilidades binomiales también se pueden hallar con el uso de varios paquetes de software. Si Y tiene una distribución binomial basada en n intentos con probabilidad de éxito p, $P(Y = y_0) = p(y_0)$ se puede hallar con el comando `dbinom(y0, n, p)`, mientras que $P(Y \leq y_0)$ se encuentra con el comando `pbinom(y0,n,p)` de *R* (o *S*-Plus). Una ventaja distintiva de usar software para calcular probabilidades binomiales es que prácticamente se pueden usar *cualesquiera* valores de n y p. Ilustramos el uso de la Tabla 1 (y, de manera simultánea, el uso de la salida del comando *R* `pbinom(y0,n,p)`) en el siguiente ejemplo.

EJEMPLO **3.9** Se supone que el voluminoso lote de fusibles eléctricos del Ejemplo 3.7 contiene sólo 5% de defectuosos. Si $n = 20$ fusibles se muestrean al azar de este lote, encuentre la probabilidad de que se observen al menos cuatro defectuosos.

Solución Si se denota con Y el número de defectuosos de la muestra, suponemos el modelo binomial para Y, con $p = .05$. Entonces,

$$P(Y \geq 4) = 1 - P(Y \leq 3),$$

y usando la Tabla 1, Apéndice 3 [o el comando `pbinom(3,20,.05)`]de R, obtenemos

$$P(Y \leq 3) = \sum_{y=0}^{3} p(y) = .984.$$

El valor .984 se encuentra en la tabla marcada $n = 20$ en la Tabla 1, Apéndice 3. Específicamente, aparece en la columna marcada $p = .05$ y en la fila marcada $a = 3$. Se deduce que

$$P(Y \geq 4) = 1 - .984 = .016.$$

Esta probabilidad es muy pequeña. Si en realidad observamos más de tres defectuosos de entre 20 fusibles, podríamos sospechar que el porcentaje reportado de 5% es erróneo. ■

La media y la varianza asociadas con una variable aleatoria binomial se deducen en el siguiente teorema. Como veremos en la prueba del teorema, es necesario evaluar la suma de algunas series aritméticas. En el curso de la prueba ilustramos algunas de las técnicas que existen para sumar esas series. En particular, usamos el hecho de que $\sum_{y} p(y) = 1$ para cualquier variable aleatoria discreta.

TEOREMA 3.7 Sea Y una variable aleatoria binomial basada en n pruebas y probabilidad p de éxito. Entonces

$$\mu = E(Y) = np \quad \text{y} \quad \sigma^2 = V(Y) = npq.$$

Demostración Por las Definiciones 3.4 y 3.7,

$$E(Y) = \sum_y yp(y) = \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y}.$$

Observe que el primer término de la suma es 0 y, por tanto, que

$$\begin{aligned} E(Y) &= \sum_{y=1}^{n} y \frac{n!}{(n-y)!y!} p^y q^{n-y} \\ &= \sum_{y=1}^{n} \frac{n!}{(n-y)!(y-1)!} p^y q^{n-y}. \end{aligned}$$

Los sumandos de esta última expresión presentan un sorprendente parecido con las probabilidades binomiales. De hecho, si factorizamos np en cada término de la suma y hacemos $z = y - 1$,

$$\begin{aligned} E(Y) &= np \sum_{y=1}^{n} \frac{(n-1)!}{(n-y)!(y-1)!} p^{y-1} q^{n-y} \\ &= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)!z!} p^z q^{n-1-z} \\ &= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^z q^{n-1-z}. \end{aligned}$$

Observe que $p(z) = \binom{n-1}{z} p^z q^{n-1-z}$ es la función de probabilidad binomial basada en $(n-1)$ pruebas. Así, $\sum_z p(z) = 1$ y se deduce que

$$\mu = E(Y) = np.$$

Del Teorema 3.6 sabemos que $\sigma^2 = V(Y) = E(Y^2) - \mu^2$. Entonces, σ^2 se puede calcular si hallamos $E(Y^2)$. Hallar $E(Y^2)$ directamente es difícil porque

$$E(Y^2) = \sum_{y=0}^{n} y^2 p(y) = \sum_{y=0}^{n} y^2 \binom{n}{y} p^y q^{n-y} = \sum_{y=0}^{n} y^2 \frac{n!}{y!(n-y)!} p^y q^{n-y}$$

y la cantidad y^2 no aparece como factor de $y!$. ¿Hacia dónde vamos ahora? Observe que

$$E[Y(Y-1)] = E(Y^2 - Y) = E(Y^2) - E(Y)$$

y, por tanto,

$$E(Y^2) = E[Y(Y-1)] + E(Y) = E[Y(Y-1)] + \mu.$$

En este caso,

$$E[Y(Y-1)] = \sum_{y=0}^{n} y(y-1)\frac{n!}{y!(n-y)!}p^y q^{n-y}.$$

Los términos primero y segundo de esta suma son iguales a cero (cuando $y = 0$ y $y = 1$). Entonces

$$E[Y(Y-1)] = \sum_{y=2}^{n} \frac{n!}{(y-2)!(n-y)!}p^y q^{n-y}.$$

(Observe la cancelación que llevó a este último resultado. La anticipación de esta cancelación es lo que en realidad motivó la consideración de $E[Y(Y-1)]$.) De nuevo, los sumandos de la última expresión se asemejan mucho a probabilidades binomiales. Factorice $n(n-1)p^2$ en cada término de la suma y sea $z = y - 2$ para obtener

$$\begin{aligned} E[Y(Y-1)] &= n(n-1)p^2 \sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!(n-y)!}p^{y-2}q^{n-y} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!}p^z q^{n-2-z} \\ &= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z} p^z q^{n-2-z}. \end{aligned}$$

De nuevo observe que $p(z) = \binom{n-2}{z}p^z q^{n-2-z}$ es la función de probabilidad binomial basada en $(n-2)$ pruebas. Así $\sum_{z=0}^{n-2} p(z) = 1$ (de nuevo usando el dispositivo ilustrado en la obtención de la media) y

$$E[Y(Y-1)] = n(n-1)p^2.$$

Entonces,

$$E(Y^2) = E[Y(Y-1)] + \mu = n(n-1)p^2 + np$$

y

$$\begin{aligned} \sigma^2 &= E(Y^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2 \\ &= np[(n-1)p + 1 - np] = np(1-p) = npq. \end{aligned}$$

Además de proporcionar las fórmulas para la media y la varianza de una variable aleatoria binomial, la deducción del Teorema 3.7 ilustra el uso de dos de los trucos más comunes, por ejemplo, usar el hecho de que $\sum p(y) = 1$ si $p(y)$ es una función de probabilidad válida y hallar $E(Y^2)$ para determinar $E[Y(Y-1)]$. Estas técnicas también serán útiles en las secciones siguientes donde consideramos otras distribuciones de probabilidad discreta y las medias y varianzas asociadas.

Una fuente de error frecuente al aplicar la distribución de probabilidad binomial a problemas prácticos es no definir cuál de los dos posibles resultados de una prueba es el exitoso. Como consecuencia de ello, q puede usarse erróneamente en lugar de p. Defina con todo cuidado un éxito y asegúrese de que p sea igual a la probabilidad de un éxito para cada aplicación.

Hasta aquí en esta sección hemos supuesto que el número de pruebas, n, y la probabilidad de éxito, p, se conocían y usamos la fórmula $p(y) = \binom{n}{y} p^y q^{n-y}$ para calcular probabilidades asociadas con variables aleatorias binomiales. En el Ejemplo 3.8 obtuvimos un valor para $P(Y \geq 9)$ y usamos esta probabilidad para llegar a una conclusión acerca de la efectividad del medicamento. El siguiente ejemplo exhibe otro uso *estadístico*, más que probabilístico, de la distribución binomial.

EJEMPLO **3.10** Suponga que se hace una encuesta a 20 personas que trabajan para una gran empresa y se les pregunta a cada uno si están a favor de la implantación de una nueva política respecto de los fondos de jubilación. Si en nuestra muestra, 6 están a favor de la nueva política, encuentre una estimación de p, la verdadera pero desconocida proporción de empleados que están a favor de la nueva política.

Solución Si Y denota el número entre los 20 que están a favor de la nueva política, es razonable concluir que Y tiene una distribución binomial con $n = 20$ para algún valor de p. Cualquiera que sea el verdadero valor de p, concluimos que la probabilidad de observar 6 de entre 20 a favor de la política es

$$P(Y = 6) = \binom{20}{6} p^6 (1 - p)^{14}.$$

Usaremos como nuestra estimación para p el valor que maximice la probabilidad de observar el valor que *realmente observamos* (6 a favor en 20 pruebas). ¿Cómo encontramos el valor de p que maximice $P(Y = 6)$?

Como $\binom{20}{6}$ es una constante (respecto de p) y $\ln(w)$ es una función creciente de w, el valor de p que maximiza $P(Y = 6) = \binom{20}{6} p^6 (1 - p)^{14}$ es igual al valor de p que maximiza $\ln[p^6(1 - p)^{14}] = [6 \ln(p) + 14 \ln(1 - p)]$.

Si evaluamos la derivada de $[6 \ln(p) + 14 \ln(1 - p)]$ con respecto a p, obtenemos

$$\frac{d[6 \ln(p) + 14 \ln(1 - p)]}{dp} = \left(\frac{6}{p}\right) - \left(\frac{14}{1 - p}\right).$$

El valor de p que maximiza (o minimiza) $[6 \ln(p) + 14 \ln(1 - p)]$[y, lo que es más importante, $p(Y = 6)$] es la solución de la ecuación

$$\frac{6}{p} - \frac{14}{1 - p} = 0.$$

Resolviendo, obtenemos $p = 6/20$.

Como la segunda derivada de $[6 \ln(p) + 14 \ln(1 - p)]$ es negativa cuando $p = 6/20$, se deduce que $[6 \ln(p) + 14 \ln(1 - p)]$[y $P(Y = 6)$] se *maximiza* cuando $p = 6/20$. Nuestra estimación para p, basada en 6 "éxitos" entre 20 pruebas es por tanto $6/20$.

La respuesta final que obtuvimos debe parecer bastante razonable. Como p es la probabilidad de un "éxito" en cualquier prueba determinada, una estimación razonable es, en realidad, la proporción de "éxitos" en nuestra muestra, en este caso $6/20$. En la siguiente sección aplicaremos esta misma técnica para obtener una estimación que no es inicialmente tan intuitiva. Como veremos en el Capítulo 9, la estimación que acabamos de obtener es la estimación de *máxima probabilidad* para p y el procedimiento empleado líneas antes es un ejemplo de la aplicación del *método de máxima probabilidad.* ■

Ejercicios

3.35 Considere la población de electores descrita en el Ejemplo 3.6. Suponga que hay $N = 5000$ electores en la población, 40% de los cuales están a favor de Jones. Identifique el evento *está a favor de Jones* como el éxito S. Es evidente que la probabilidad de S en el intento 1 es .40. Considere el evento B de que S suceda en la segunda prueba. Entonces B puede ocurrir en dos formas: las primeras dos pruebas son exitosas *o bien* la primera prueba es un fracaso y la segunda es un éxito. Demuestre que $P(B) = .4$. ¿Cuál es $P(B|\text{la primera prueba es } S)$? ¿Esta probabilidad *condicional* difiere marcadamente de $P(B)$?

3.36 **a** Un meteorólogo en Denver registró $Y =$ el número de días de lluvia durante un periodo de 30 días. ¿Tiene Y una distribución binomial? Si es así, ¿se dan los valores de n y p?

b Una empresa de investigación de mercado ha contratado operadoras para que realicen encuestas por teléfono. Se usa una computadora para marcar al azar un número telefónico y la operadora pregunta a la persona que contesta si tiene tiempo para responder unas preguntas. Sea $Y =$ el número de llamadas hechas hasta que la primera persona contesta que está dispuesta a responder las preguntas. ¿Es éste un experimento binomial? Explique.

3.37 En 2003, el promedio de calificación combinada del examen Scholastic Aptitude Test (SAT) (matemáticas y verbal) para estudiantes que van a la universidad en Estados Unidos fue 1026. Suponga que aproximadamente 45% de todos los graduados de preparatoria hizo este examen y que 100 egresados de preparatoria se seleccionan al azar de entre todos los egresados en Estados Unidos. De las siguientes variables aleatorias, ¿cuál tiene una distribución que puede ser aproximada por una distribución binomial? Siempre que sea posible, dé los valores para n y p.

a El número de estudiantes que hizo el SAT
b Las calificaciones de los 100 estudiantes de la muestra
c El número de estudiantes de la muestra que obtuvo calificaciones arriba del promedio del SAT
d El tiempo necesario para que cada estudiante terminara el SAT
e El número de egresadas (mujeres) de preparatoria de la muestra

3.38 El fabricante de una bebida láctea de bajo contenido de calorías desea comparar el atractivo del gusto de una nueva fórmula (fórmula B) con el de la fórmula estándar (fórmula A). A cada uno de cuatro jueces se les dan tres vasos en orden aleatorio, dos de ellos cont la fórmula A y el otro con la fórmula B. A cada uno de los jueces se les pide indicar cuál vaso fue el que disfrutó más. Suponga que las dos fórmulas son igualmente atractivas. Sea Y el número de jueces que indican una preferencia por la nueva fórmula.

a Encuentre la función de probabilidad para Y.
b ¿Cuál es la probabilidad de que al menos tres de los cuatro jueces indique una preferencia por la nueva fórmula?
c Encuentre el valor esperado de Y.
d Encuentre la varianza de Y.

3.39 Se construye un complejo sistema electrónico con cierto número de piezas de respaldo en sus subsistemas. Un subsistema tiene cuatro componentes idénticos, cada uno con una probabilidad de .2 de fallar en menos de 1000 horas. El subsistema va a operar si dos de los cuatro componentes están operando. Suponga que los componentes operan de manera independiente. Encuentre la probabilidad de que

a exactamente dos de los cuatro componentes dure más de 1000 horas,
b el subsistema opere más de 1000 horas.

3.40 La probabilidad de que un paciente se recupere de una enfermedad estomacal es .8. Suponga que se sabe que 20 personas han contraído la enfermedad. ¿Cuál es la probabilidad de que

a exactamente 14 se recuperen?,
b al menos 10 se recuperen?,
c al menos 14 pero no más de 18 se recuperen?,
d a lo sumo 16 se recuperen?

3.41 Un examen de opción múltiple tiene 15 preguntas, cada una con cinco posibles respuestas, sólo una de las cuales es correcta. Suponga que uno de los estudiantes que hace el examen contesta cada una de las preguntas con una adivinación aleatoria independiente. ¿Cuál es la probabilidad de que conteste correctamente al menos diez preguntas?

3.42 Consulte el Ejercicio 3.41. ¿Cuál es la probabilidad de que un estudiante conteste correctamente al menos diez preguntas si

a para cada pregunta, el estudiante puede eliminar correctamente una de las respuestas equivocadas y contesta a continuación cada una de las preguntas con una adivinación aleatoria independiente entre las respuestas restantes?,
b puede eliminar correctamente dos respuestas equivocadas para cada pregunta y al azar selecciona de entre las respuestas restantes?

3.43 Muchas empresas que generan energía eléctrica promueven el ahorro de energía, para lo cual ofrecen tarifas de descuento a consumidores que mantengan su uso por debajo de ciertos estándares establecidos de subsidio. Un reciente informe de la EPA (agencia de protección ambiental observa que 70% de los residentes en Puerto Rico han reducido su consumo de electricidad lo suficiente para tener derecho a tarifas de descuento. Si se seleccionan al azar cinco suscriptores residenciales de San Juan, Puerto Rico, encuentre la probabilidad de cada uno de los siguientes eventos:

a Los cinco tienen derecho a las tarifas favorables,
b Al menos cuatro tienen derecho a tarifas favorables,

3.44 Un nuevo procedimiento quirúrgico es exitoso con una probabilidad de p. Suponga que la operación se realiza cinco veces y los resultados son independientes entre sí. ¿Cuál es la probabilidad de que

a todas las operaciones sean exitosas si $p = .8$?
b exactamente cuatro sean exitosas si $p = .6$?
c menos de dos sean exitosas si $p = .3$?

3.45 Un aparato detector de incendios utiliza tres celdas sensibles a la temperatura que actúan de manera independiente una de otra, en forma tal que una o más pueden activar la alarma. Cada celda presenta una probabilidad de $p = .8$ de activar la alarma cuando la temperatura alcanza 100°C o más. Sea Y igual al número de celdas que activan la alarma cuando la temperatura alcanza los 100°.

a Encuentre la distribución de probabilidad para Y.
b Encuentre la probabilidad de que la alarma funcione cuando la temperatura alcance los 100°.

3.46 Construya histogramas de probabilidad para las distribuciones de probabilidad binomiales para $n = .5$, $p = .1$, .5 y .9. (La Tabla 1, Apéndice 3, reducirá la cantidad de cálculos.) Nótese la simetría para $p = .5$ y la dirección del sesgo para $p = .1$ y .9.

3.47 Use la Tabla 1, Apéndice 3, para construir un histograma de probabilidad para la distribución de probabilidad binomial para $n = 20$ y $p = .5$. Observe que casi toda la probabilidad cae en el intervalo $5 \le y \le 15$.

3.48 Un sistema de protección contra proyectiles está formado por n aparatos de radar que operan de manera independiente, cada uno con una probabilidad de .9 de detectar un proyectil que penetre en una zona que está cubierta por todas las unidades.

a Si $n = 5$ y un proyectil entra en la zona, ¿cuál es la probabilidad de que exactamente cuatro aparatos lo detecten? ¿Al menos un aparato?

b ¿Qué tan grande debe ser n si requerimos que la probabilidad de detectar un proyectil que entre en la zona sea .999?

3.49 Un fabricante de cera para pisos ha creado dos nuevas marcas, A y B, que desea someter a evaluación de propietarios de casas para determinar cuál de las dos es superior. Ambas ceras, A y B, se aplican a superficies de pisos en cada una de 15 casas. Suponga que en realidad no hay diferencia en la calidad de las marcas. ¿Cuál es la probabilidad de que diez o más propietarios de casas expresen preferencia por

a la marca A?,

b ya sea la marca A o la marca B?

3.50 En el Ejercicio 2.151 consideramos un modelo para la Serie Mundial de Beisbol. Dos equipos A y B juegan una serie de partidos hasta que un equipo gane cuatro de ellos. Suponemos que los partidos se juegan de manera independiente y que la probabilidad de que A gane cualquier partido es p. Calcule la probabilidad de que la serie dure exactamente cinco partidos. [*Sugerencia:* use lo que sepa acerca de la variable aleatoria, Y, el número de partidos que A gane entre los primeros cuatro partidos.]

3.51 En el siglo XVIII, Chevalier de Mere pidió a Blaise Pascal comparar las probabilidades de dos eventos. A continuación, usted va a calcular la probabilidad de los dos eventos que, antes de una experiencia contraria en juegos, eran considerados por de Mere como igualmente probables.

a ¿Cuál es la probabilidad de obtener al menos un 6 en cuatro tiros de un dado sin cargar?

b Si un par de dados sin cargar se lanza 24 veces en una mesa, ¿cuál es la probabilidad de que salga al menos un doble seis?

3.52 La prueba de gusto para la PTC (feniltiocarbamida) es un ejercicio favorito en grupos de estudiantes principiantes de genética humana. Se ha establecido que un solo gen determina si un individuo es un "probador" o no lo es. Si 70% de norteamericanos son "probadores" y 20 de ellos se seleccionan al azar, ¿cuál es la probabilidad de que

a al menos 17 sean "probadores"?,

b menos de 15 sean "probadores"?

3.53 La enfermedad de Tay-Sachs es una afección genética que suele ser mortal en niños. Si ambos padres son portadores de la enfermedad, la probabilidad de que sus hijos desarrollen la enfermedad es aproximadamente .25. Suponga que un esposo y esposa son portadores y que tienen tres hijos. Si los resultados de los tres embarazos son mutuamente independientes, ¿cuáles son las probabilidades de los siguientes eventos?

a Los tres hijos desarrollan la enfermedad.

b Sólo uno de los hijos desarrolla la enfermedad.

c El tercer hijo desarrolla la enfermedad, dado que los primeros dos no la desarrollaron.

3.54 Suponga que Y es una variable aleatoria binomial basada en n intentos con probabilidad p de éxito y considere $Y^* = n - Y$.

a Demuestre que para $y^* = 0, 2, \ldots, n$

$$P(Y^* = y\ ) = P(n - Y^* = y^*) = P(Y = n - Y^*).$$

b Use el resultado del inciso a para demostrar que

$$P(Y^* = y^*) = \binom{n}{n - y^*} p^{n-y^*} q^{y^*} = \binom{n}{y^*} q^{y^*} p^{n-y^*}.$$

c El resultado del inciso b implica que Y^* tiene una distribución binomial basada en n pruebas y probabilidad de "éxito" $p^* = q = 1 - p$. ¿Por qué este resultado es "obvio"?

3.55 Suponga que Y es una variable aleatoria binomial con $n > 2$ pruebas y una probabilidad p de éxito. Use la técnica presentada en el Teorema 3.7 y el hecho de que $E\{Y(Y-1)(Y-2)\} = E(Y^3) - 3E(Y^2) + 2E(Y)$ para deducir $E(Y^3)$.

3.56 Una empresa de exploración petrolera se forma con suficiente capital para financiar diez exploraciones. La probabilidad de que una exploración particular sea exitosa es .1. Suponga que las exploraciones son independientes. Encuentre la media y la varianza del número de exploraciones exitosas.

3.57 Consulte el Ejercicio 3.56. Suponga que la empresa tiene un costo fijo de $20,000 por preparar equipo antes de hacer su primera exploración. Si cada exploración exitosa cuesta $30,000 y cada una no exitosa cuesta $15,000, encuentre el costo esperado total para la empresa por sus diez exploraciones.

3.58 Una venta particular comprende cuatro artículos seleccionados al azar de un lote voluminoso que se sabe contiene 10% de artículos defectuosos. Denote con Y el número de los defectuosos entre los cuatro vendidos. El comprador de los artículos devolverá los defectuosos para su reparación y el costo de ésta está dado por $C = 3Y^2 + Y + 2$. Encuentre el costo de reparación esperado. [*Sugerencia:* el resultado del Teorema 3.6 implica que, para cualquier variable aleatoria Y, $E(Y^2) = \sigma^2 + \mu^2$.]

3.59 Diez motores se empacan para su venta en cierto almacén. Los motores se venden en $100 cada uno, pero una garantía de devolución del doble de su dinero es efectiva por cualquier unidad defectuosa que el comprador pueda recibir. Encuentre la ganancia neta esperada para el vendedor si la probabilidad de que cualquier motor sea defectuoso es .08. (Suponga que la calidad de cualquier motor es independiente de la de los otros.)

3.60 Una concentración particular de un producto químico detectado en agua contaminada se encuentra que es letal para 20% de los peces que queden expuestos a la concentración durante 24 horas. Veinte peces se colocan en un tanque que contiene esta concentración del producto químico en agua.

a Encuentre la probabilidad de que exactamente 14 sobrevivan.
b Encuentre la probabilidad de que al menos 10 sobrevivan.
c Encuentre la probabilidad de que a lo sumo 16 sobrevivan.
d Encuentre la media y la varianza del número que sobrevive.

3.61 De los donadores voluntarios de sangre en una clínica, 80% presentan factor Rhesus (Rh) en su sangre.

a Si cinco voluntarios se seleccionan al azar, ¿cuál es la probabilidad de que al menos uno no presente el factor Rh?

b Si cinco voluntarios se seleccionan al azar, ¿cuál es la probabilidad de que a lo sumo cuatro presenten el factor Rh?

c ¿Cuál es el número más pequeño de voluntarios que deben seleccionarse si deseamos tener certeza de al menos 90% de obtener al menos cinco donadores con factor Rh?

3.62 Goranson y Hall (1980) explican que la probabilidad de detectar una grieta en el ala de un avión es el producto de p_1, la probabilidad de inspeccionar un avión con una grieta en un ala; p_2, la probabilidad de inspeccionar el detalle donde se ubica la grieta; y p_3, la probabilidad de detectar el daño.

a ¿Qué suposiciones justifican la multiplicación de estas probabilidades?

b Suponga que $p_1 = .9$, $p_2 = .8$ y $p_3 = .5$ para cierta flota de aviones. Si se inspeccionan tres aviones de esta flota, encuentre la probabilidad de que una grieta en el ala sea detectada en al menos uno de ellos.

***3.63** Considere la distribución binomial con n pruebas y $P(S) = p$.

a Demuestre que $\frac{p(y)}{p(y-1)} = \frac{(n-y+1)p}{yq}$ para $y = 1, 2, \ldots, n$. De modo equivalente, para $y = 1, 2, \ldots, n$, la ecuación $p(y) = \frac{(n-y+1)p}{yq}p(y-1)$ da una relación repetitiva entre las probabilidades asociadas con valores sucesivos de Y.

b Si $n = 90$ y $p = .04$, use la relación citada en el inciso anterior para hallar $P(Y < 3)$.

c Demuestre que $\frac{p(y)}{p(y-1)} = \frac{(n-y+1)p}{yq} > 1$ si $y < (n+1)p$, que $\frac{p(y)}{p(y-1)} < 1$ si $y > (n+1)p$, y que $\frac{p(y)}{p(y-1)} = 1$ si $(n+1)$ es un entero y $y = (n+1)p$. Esto establece que $p(y) > p(y-1)$ si y es pequeña ($y < (n+1)p$) y $p(y) < p(y-1)$ si y es grande ($y > (n+1)p$). Entonces, las probabilidades binomiales sucesivas aumentan por algún tiempo y disminuyen de ahí en adelante.

d Demuestre que el valor de y asignado a la probabilidad más grande es igual al máximo entero menor o igual que $(n+1)p$. Si $(n+1)p = m$ para algún entero m, entonces $p(m) = p(m-1)$.

***3.64** Considere una extensión de la situación estudiada en el Ejemplo 3.10. Si hay n pruebas en un experimento binomial y observamos y_0 "éxitos", demuestre que $P(Y = y_0)$ se maximiza cuando $p = y_0/n$. De nuevo, estamos determinando (en general esta vez) el valor de p que maximice la probabilidad del valor de Y que observamos realmente.

***3.65** Consulte el Ejercicio 3.64. El *estimador de la máxima probabilidad* para p es Y/n (nótese que Y es la variable aleatoria binomial, no un valor particular de ella).

a Deduzca $E(Y/n)$. En el Capítulo 9 veremos que este resultado implica que Y/h es un estimador *no sesgado* para p.

b Deduzca $V(Y/n)$. ¿Qué le ocurre a $V(Y/n)$ cuando n se hace grande?

3.5 La distribución de probabilidad geométrica

La variable aleatoria con distribución de probabilidad geométrica está asociada con un experimento que comparte algunas de las características de un experimento binomial. Este experimento también comprende pruebas idénticas e independientes, cada una de las cuales puede arrojar uno de dos resultados: éxito o fracaso. La probabilidad de éxito es igual a p y es constante de una prueba a otra. No obstante, en lugar del número de éxitos que se presentan en n pruebas, la variable aleatoria geométrica Y es el número de prueba en la que ocurre el primer éxito. Entonces, el experimento consiste en una serie de pruebas que concluye con el primer éxito. En consecuencia, el experimento podría terminar con la primera prueba si se observa un éxito en la misma o el experimento podría continuar de manera indefinida.

El espacio muestral S para el experimento contiene el conjunto infinitamente contable de puntos muestrales:

E_1: S (éxito en la primera prueba)
E_2: FS (fracaso en la primera, éxito en la segunda)
E_3: FFS (primer éxito en la tercera prueba)
E_4: $FFFS$ (primer éxito en la cuarta prueba)
⋮
E_k: $\underbrace{FFFF\ldots F}_{k-1}S$ (primer éxito en la $k^{\text{ésima}}$ prueba)
⋮

Como la variable aleatoria Y es el número de intentos hasta el primer éxito, incluido éste los eventos $(Y = 1)$, $(Y = 2)$ y $(Y = 3)$ contienen sólo los puntos muestrales E_1, E_2 y E_3, respectivamente. En forma más general, el evento numérico $(Y = y)$ contiene sólo a E_y. Como las pruebas son independientes, para cualquier $y = 1, 2, 3, \ldots,$

$$p(y) = P(Y = y) = P(E_y) = P(\underbrace{FFFF\ldots F}_{y-1}S) = \underbrace{qqq\cdots q}_{y-1}\,p = q^{y-1}p.$$

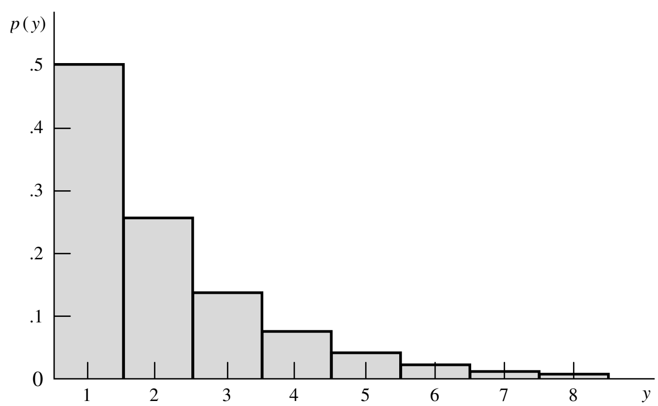
DEFINICIÓN **3.8**

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad geométrica* si y sólo si

$$p(y) = q^{y-1}p, \qquad y = 1,\ 2,\ 3, \ldots,\quad 0 \le p \le 1.$$

En la Figura 3.5 se ilustra un histograma de probabilidad para $p(y)$, $p = .5$. Las áreas sobre los intervalos corresponden a probabilidades, como correspondieron a las distribuciones de frecuencia de datos en el Capítulo 1, excepto que Y puede tomar sólo valores discretos, $y = 1, 2, \ldots, \infty$. Por inspección de los valores respectivos es obvio que $p(y) \geq 0$. En el Ejercicio 3.66 demostrará que estas probabilidades ascienden a 1, como se requiere para cualquier distribución de probabilidad discreta válida.

FIGURA **3.5** La distribución de probabilidad geométrica, $p = .5$



La distribución de probabilidad geométrica se emplea con frecuencia para modelar distribuciones de la duración de tiempos de espera. Por ejemplo, suponga que el motor de un avión comercial recibe atención periódicamente para cambiar sus diversas partes en puntos diferentes de tiempo y que, por tanto, son de edades que varían. Entonces la probabilidad p de mal funcionamiento del motor durante cualquier intervalo de operación de una hora observado al azar podría ser el mismo que para cualquier otro intervalo de una hora. El tiempo transcurrido antes de que el motor falle es el número de intervalos de una hora, Y, hasta que ocurra la primera falla. (Para esta aplicación, el mal funcionamiento del motor en un periodo determinado de una hora se define como éxito. Observe que como en el caso del experimento binomial, cualquiera de los dos resultados de una prueba se puede definir como éxito. De nuevo, un "éxito" no es necesariamente lo que podría considerarse como "bueno" en nuestra conversación de todos los días.)

EJEMPLO **3.11** Suponga que la probabilidad de mal funcionamiento de un motor durante cualquier periodo de una hora es $p = .02$. Encuentre la probabilidad de que un motor determinado funcione bien dos horas.

Solución Si denotamos con Y el número de intervalos de una hora hasta la primera falla, tenemos

$$P(\text{funciona bien dos horas}) = P(Y \geq 3) = \sum_{y=3}^{\infty} p(y).$$

Como $\sum_{y=1}^{\infty} p(y) = 1$,

$$\begin{aligned} P(\text{funciona bien dos horas}) &= 1 - \sum_{y=1}^{2} p(y) \\ &= 1 - p - qp = 1 - .02 - (.98)(.02) = .9604. \end{aligned}$$

■

Si examina la fórmula para la distribución geométrica dada en la Definición 3.8, verá que valores más grandes de p (y por tanto valores menores de q) llevan a probabilidades más altas para los valores más pequeños de Y y en consecuencia a menores probabilidades para los valores más grandes de Y. Así, el valor medio de Y parece ser inversamente proporcional a p. Como demostramos en el siguiente teorema, la media de una variable aleatoria con distribución geométrica en realidad es igual a $1/p$.

TEOREMA **3.8** Si Y es una variable aleatoria con una distribución geométrica,

$$\mu = E(Y) = \frac{1}{p} \quad \text{y} \quad \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Demostración

$$E(Y) = \sum_{y=1}^{\infty} yq^{y-1}p = p\sum_{y=1}^{\infty} yq^{y-1}.$$

Esta serie podría parecer difícil de sumarse directamente. En realidad, se puede sumar con facilidad si tomamos en cuenta que, para $y \geq 1$,

$$\frac{d}{dq}(q^y) = yq^{y-1},$$

y, por tanto,

$$\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right) = \sum_{y=1}^{\infty} yq^{y-1}.$$

(El intercambio de derivada y suma aquí se puede justificar.) Sustituyendo, obtenemos

$$E(Y) = p\sum_{y=1}^{\infty} yq^{y-1} = p\,\frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right).$$

La última suma es la serie geométrica, $q + q^2 + q^3 + \ldots$, que es igual a $q/(1-q)$(vea el Apéndice $A1.11$). Por tanto,

$$E(Y) = p\,\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left[\frac{1}{(1-q)^2}\right] = \frac{p}{p^2} = \frac{1}{p}.$$

Para resumir, nuestro método es expresar una serie que no se puede sumar directamente como la derivada de una serie para la cual la suma se puede obtener con facilidad. Una vez que evaluemos la serie más fácilmente manejable, derivamos para completar el proceso.

La derivación de la varianza se deja como Ejercicio 3.85.

EJEMPLO **3.12** Si la probabilidad de falla de un motor durante cualquier periodo de una hora es $p = .02$ y Y denota el número de intervalos de una hora hasta la primera falla, encuentre la media y la desviación estándar de Y.

Solución Al igual que en el Ejemplo 3.11, se deduce que Y tiene una distribución geométrica con $p = .02$. Entonces, $E(Y) = 1/p = 1/(.02) = 50$, y esperaríamos unas pocas horas antes de encontrar una falla. Además, $V(Y) = .98/.0004 = 2450$, y se deduce que la desviación estándar de Y es $\sigma = \sqrt{2450} = 49.497$. ■

Aun cuando el cálculo de probabilidades asociado con variables aleatorias geométricas se puede efectuar al evaluar un solo valor o sumas parciales asociadas con una serie geométrica, estas probabilidades también se pueden hallar con el uso de varios paquetes de software. Si Y tiene una distribución geométrica con probabilidad p de éxito, $P(Y = y_0) = p(y_0)$ se puede hallar con el uso del comando `dgeom(y0-1,p)` de R (o S-Plus), mientras que $P(Y \leq y_0)$ se encuentra usando el comando `pgeom(y0-1,p)` de R (o S-Plus). Por ejemplo el comando `pgeom(1,0.02)` de *R* (o *S*-Plus) da el valor para $P(Y \leq 2)$ que implícitamente se usó en el Ejemplo 3.11. Observe que el argumento en estos comandos es el valor $y_0 - 1$, no el valor y_0. Esto es porque algunos autores prefieren definir que la distribución geométrica sea la de la variable aleatoria $Y^* =$ *el número de fracasos antes del primer éxito.* En nuestra formulación, la variable aleatoria geométrica Y se interpreta como *el número de intento en los que ocurre el primer éxito.* En el Ejercicio 3.88 verá que $Y^* = Y - 1$. Debido a esta relación entre las dos versiones de variables aleatorias geométricas, $P(Y = y_0) = P(Y - 1 = y_0 - 1) = P(Y^\star = y_0 - 1)$. *R* calcula probabilidades asociadas con Y^*, explicando por qué los argumentos para `dgeom` y `pgeom` son $y_0 - 1$ en lugar de y_0.

El siguiente ejemplo, semejante al Ejemplo 3.10, ilustra la forma en que la distribución de probabilidad geométrica se puede usar para estimar un valor desconocido de p, la probabilidad de un éxito.

EJEMPLO **3.13** Suponga que entrevistamos personas sucesivas que trabajan para la gran empresa estudiada en el Ejemplo 3.10 y detenemos las entrevistas cuando encontramos a la primera persona que le guste esa política. Si la quinta persona entrevistada es la primera que está a favor de la nueva política, encuentre una estimación para p, la proporción verdadera pero desconocida de empleados que están a favor de la nueva política.

Solución Si Y denota el número de personas entrevistadas hasta que hallemos la primera a quien le guste el nuevo plan de retiro, es razonable concluir que Y tiene una distribución geométrica para algún valor de p. Cualquiera que sea el verdadero valor para p, concluimos que la probabilidad de observar la primera persona a favor de la política en el quinto intento es

$$P(Y = 5) = (1 - p)^4 p.$$

Usaremos como nuestra estimación para p, el valor que maximice la probabilidad de observar el valor que *en realidad observamos* (el primer éxito en el intento 5).

Para hallar el valor de p que maximice a $P(Y = 5)$, de nuevo observamos que el valor de p que maximice a $P(Y = 5) = (1 - p)^4 p$ es igual que el valor de p que maximice a $\ln[(1 - p)^4 p] = [4 \ln(1 - p) + \ln(p)]$.

Si evaluamos la derivada de $[4 \ln(1 - p) + \ln(p)]$ con respecto a p, obtenemos

$$\frac{d[4 \ln(1 - p) + \ln(p)]}{dp} = \frac{-4}{1 - p} + \frac{1}{p}.$$

Si igualamos a cero esta derivada y resolvemos, obtenemos $p = 1/5$.

Como la segunda derivada de $[4 \ln(1 - p) + \ln(p)]$ es negativa cuando $p = 1/5$, se deduce que $[4 \ln(1 - p) + \ln(p)]$[y $P(Y = 5)$] se *maximiza* cuando $p = 1/5$. Nuestra estimación para p, basada en observar el primer éxito en el quinto intento es $1/5$.

Quizá este resultado es un poco más sorprendente que la respuesta que obtuvimos en el Ejemplo 3.10 cuando estimamos p con base en observar 6 a favor del nuevo plan en una muestra de tamaño 20. De nuevo, éste es un ejemplo del uso del *método de máxima probabilidad* que estudiaremos con más detalle en el Capítulo 9. ■

Ejercicios

3.66 Suponga que Y es una variable aleatoria con distribución geométrica. Demuestre que

a $\sum_y p(y) = \sum_{y=1}^{\infty} q^{y-1} p = 1$.

b $\dfrac{p(y)}{p(y-1)} = q$, para $y = 2, 3, \ldots$. Esta relación es menor que 1, lo cual implica que las probabilidades geométricas están decreciendo en forma monotónica como función de y. Si Y tiene una distribución geométrica, ¿qué valor de Y es el más probable (tiene la más alta probabilidad)?

3.67 Suponga que 30% de los solicitantes para cierto trabajo industrial posee capacitación avanzada en programación computacional. Los candidatos son elegidos aleatoriamente entre la población y entrevistados en forma sucesiva. Encuentre la probabilidad de que el primer solicitante con capacitación avanzada en programación se encuentre en la quinta entrevista.

3.68 Consulte el Ejercicio 3.67. ¿Cuál es el número esperado de solicitantes que será necesario entrevistar para hallar el primero con capacitación avanzada?

3.69 Unos seis meses después del segundo periodo de George W. Bush como presidente, una encuesta de Gallup indicó que un nivel (bajo) muy cerca del récord de 41% de adultos expresaron "mucha" o "bastante" confianza en la suprema corte de Estados Unidos (http://www.gallup.com./poll/content/default.aspx?ci=17011), junio de 2005). Supongamos que usted realizó su propia encuesta telefónica en ese tiempo y al azar llamó a personas y les pidió describieran su nivel de confianza en la suprema corte. Encuentre la distribución de probabilidad de Y, el número de llamadas hasta que se encuentre la primera persona que *no* exprese "mucha" o "bastante" confianza en la suprema corte de Estados Unidos.

3.70 Una empresa de exploración petrolera va a hacer una serie de perforaciones de sondeo en una zona determinada en busca de un pozo productivo. La probabilidad de que tenga éxito en un intento dado es .2.

a ¿Cuál es la probabilidad de que la tercera perforación sea la primera en dar un pozo productivo?

b Si la empresa puede darse el lujo de perforar a lo sumo diez pozos, ¿cuál es la probabilidad de que no encuentre un pozo productivo?

3.71 Denote con Y una variable aleatoria geométrica con probabilidad de éxito p.

a Demuestre que para un entero positivo a,

$$P(Y > a) = q^a.$$

b Demuestre que para los enteros positivos a y b,

$$P(Y > a + b | Y > a) = q^b = P(Y > b).$$

Este resultado implica que, por ejemplo, $P(Y > 7 | Y > 2) = P(Y > 5)$. ¿Por qué supone que esta propiedad recibe el nombre de propiedad *sin memoria* de la distribución geométrica?

c En el desarrollo de la distribución de la variable aleatoria geométrica supusimos que el experimento consistió en realizar intentos idénticos e independientes hasta que se observó el primer éxito. En vista de estas suposiciones, ¿por qué es "obvio" el resultado en el inciso b?

3.72 Dado que ya hemos lanzado al aire una moneda balanceada diez veces y no obtuvimos caras, ¿cuál es la probabilidad de que debemos lanzarla al menos dos veces más para obtener la primera cara?

3.73 Un contador público certificado (CPA, por sus siglas en inglés) ha encontrado que nueve de entre diez compañías auditadas contienen errores importantes. Si el CPA hace auditoría a una serie de cuentas de empresas, ¿cuál es la probabilidad de que la primera cuenta que contenga errores importantes

a sea la tercera en ser auditada?,

b sea la tercera cuenta auditada la que le sigue?

3.74 Consulte el Ejercicio 3.73. ¿Cuáles son la media y la desviación estándar del número de cuentas que deben ser examinadas para hallar la primera con errores importantes?

3.75 La probabilidad de que llegue un cliente al mostrador de servicio de una tienda en un segundo cualquiera es igual a .1. Suponga que llegan clientes en forma aleatoria y por tanto que una llegada en un segundo cualquiera es independiente de las otras. Encuentre la probabilidad de que la primera llegada

a ocurra durante el tercer intervalo de un segundo.

b no ocurra hasta al menos el tercer intervalo de un segundo.

3.76 Si Y tiene una distribución geométrica con probabilidad de éxito .3, ¿cuál es el máximo valor, y_0, tal que $P(Y > y_0) \geq .1$?

3.77 Si Y tiene una distribución geométrica con probabilidad p de éxito, demuestre que

$$P(Y = \text{un entero impar}) = \frac{p}{1 - q^2}.$$

3.78 De una población de consumidores, 60% tienen fama de preferir una marca particular, A, de pasta dental. Si se entrevista a un grupo de consumidores escogidos al azar, ¿cuál es la probabilidad de que exactamente cinco personas tengan que ser entrevistadas para hallar el primer consumidor que prefiera la marca A? ¿Al menos cinco personas?

3.79 Al contestar una pregunta de encuesta en un tema delicado (por ejemplo: "¿Ha fumado mariguana alguna vez?"), muchas personas prefieren no contestar de manera afirmativa. Suponga que 80% de la población no ha fumado mariguana y que todos contestan negativamente con verdad a la pregunta. El 20% restante de la población han fumado mariguana y 70% de ellos mentirán. Deduzca la distribución de probabilidad de Y, el número de personas a las que sería necesario preguntar para obtener una sola respuesta afirmativa.

3.80 Dos personas, por turnos, tiran un dado imparcial hasta que una de ellas lanza un 6. La persona A tiró primero, la B en segundo, A en tercero y así sucesivamente. En vista de que la persona B tiró el primer 6, ¿cuál es la probabilidad de que B obtenga el primer 6 en su segundo tiro (es decir, en el cuarto tiro total)?

3.81 ¿Cuántas veces esperaría usted lanzar al aire una moneda balanceada para obtener la primera cara?

3.82 Consulte el Ejercicio 3.70. La empresa hace perforaciones de exploración hasta que encuentra un pozo productivo. ¿Cuántas perforaciones esperaría hacer la empresa? Interprete intuitivamente su respuesta.

3.83 Al secretario de los Ejercicios 2.121 y 3.16 se le dieron n contraseñas de computadora y la prueba al azar. Exactamente una de las contraseñas permite el acceso a un archivo de computadora. Suponga ahora que el secretario selecciona una contraseña, la intenta y si no funciona, la regresa con las otras antes de seleccionar al azar la siguiente (¡no es muy buen secretario!) ¿Cuál es la probabilidad de hallar la contraseña correcta en el sexto intento?

3.84 Consulte el Ejercicio 3.83. Encuentre la media y la varianza de Y, el número de intento en el que se identifica la contraseña correcta.

***3.85** Encuentre $E[Y(Y-1)]$ para una variable aleatoria geométrica Y al hallar $d^2/dq^2\left(\sum_{y=1}^{\infty} q^y\right)$. Use este resultado para hallar la varianza de Y.

***3.86** Considere una extensión de la situación examinada en el Ejemplo 3.13. Si observamos y_0 como el valor para una variable aleatoria geométrica Y, demuestre que $P(Y = y_0)$ se maximiza cuando $p = 1/y_0$. De nuevo, estamos determinando (en general esta vez) el valor de p que maximiza la probabilidad del valor de Y que en realidad observamos.

***3.87** Consulte el Ejercicio 3.86. El *estimador de la máxima probabilidad* para p es $1/Y$ (observe que Y es la variable aleatoria geométrica, no un valor particular de ella). Deduzca $E(1/Y)$. [*Sugerencia:* si $|r| < 1$, $\sum_{i=1}^{\infty} r^i/i = -\ln(1-r)$.]

***3.88** Si Y es una variable aleatoria geométrica, defina $Y^* = Y-1$. Si Y se interpreta como el número del intento en el que ocurre el primer éxito, entonces Y^* se puede interpretar como el número de fracasos *antes* del primer éxito. Si $Y^* = Y-1$, $P(Y^* = y) = P(Y-1 = y) = P(Y = y+1)$ para $y = 0, 1, 2, \ldots$ Demuestre que

$$P(Y^* = y) = q^y p, \qquad y = 0, 1, 2, \ldots .$$

La distribución de probabilidad de Y^* es empleada en ocasiones por actuarios como modelo para la distribución del número de reclamaciones de seguro hechas en un tiempo específico.

***3.89** Consulte el Ejercicio 3.88. Deduzca la media y la varianza de la variable aleatoria Y^*

a usando el resultado del Ejercicio 3.33 y la relación $Y^* = Y-1$, donde Y es geométrica,

***b** directamente, usando la distribución de probabilidad para Y^* dada en el Ejercicio 3.88.

3.6 La distribución de probabilidad binomial negativa (opcional)

Una variable aleatoria con distribución binomial negativa se origina de un contexto semejante al que da la distribución geométrica. De nuevo nos concentramos en intentos independientes e idénticos, cada uno de los cuales conduce a uno de dos resultados: éxito o fracaso. La probabilidad p de éxito sigue siendo igual de un intento a otro. La distribución geométrica maneja el caso donde estamos interesados en el número de intento en el que ocurre el primer éxito. ¿Qué pasa si estamos interesados en conocer el número de intento en el que ocurre el éxito segundo, tercero o cuarto? La distribución que se aplica a la variable aleatoria Y igual al número del intento en el que ocurre el r–ésimo éxito ($r = 2, 3, 4$, etc.) es la distribución binomial negativa.

Los pasos siguientes se asemejan estrechamente a los de la sección anterior. Seleccionemos valores fijos para r y y y consideremos los eventos A y B, donde

$$A = \{\text{los primeros } (y-1) \text{ intentos contienen } (r-1) \text{ éxitos}\}$$

y

$$B = \{\text{el intento } y \text{ resulta en un éxito}\}.$$

Como suponemos que los intentos son independientes, se deduce que A y B son eventos independientes y las suposiciones previas implican que $P(B) = p$. Por tanto,

$$p(y) = p(Y = y) = P(A \cap B) = P(A) \times P(B).$$

Observe que $P(A)$ es 0 si $(y-1) < (r-1)$ o, de igual manera, si $y < r$. Si $y \geq r$, nuestro trabajo previo con la distribución binominal implica que

$$P(A) = \binom{y-1}{r-1} p^{r-1} q^{y-r}.$$

Por último,

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r,\ r+1,\ r+2, \ldots$$

DEFINICIÓN 3.9

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad binomial negativa* si y sólo si

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r,\ r+1,\ r+2, \ldots,\quad 0 \le p \le 1.$$

EJEMPLO 3.14 Un estudio geológico indica que un pozo petrolero de exploración perforado en una región particular debe producir petróleo con probabilidad .2. Encuentre la probabilidad de que el tercer descubrimiento de petróleo llegue en el quinto pozo perforado.

Solución Suponiendo perforaciones independientes y probabilidad .2 de descubrir petróleo en cualquiera de los pozos, denote con Y el número del intento en el que ocurre el tercer descubrimiento de petróleo. Entonces, es razonable suponer que Y tiene una distribución binomial negativa con $p = .2$. Como estamos interesados en $r = 3$ y $y = 5$,

$$\begin{aligned} P(Y = 5) = p(5) &= \binom{4}{2}(.2)^3(.8)^2 \\ &= 6(.008)(.64) = .0307. \end{aligned}$$

■

Si $r = 2, 3, 4, \ldots$ y Y tiene una distribución binomial negativa con probabilidad p de éxito, $P(Y = y_0) = p(y_0)$ se puede hallar usando el comando `pnbinom(y0-r,r,p)` de R (o S-Plus). Si queremos utilizar R para obtener $p(5)$ en el Ejemplo 3.14, usamos el comando `dnbinom(2,3,.2)`. Alternativamente, $P(Y \le y_0)$ se encuentra usando el comando `pnbinom(y0-r,r,p)` de R (o S-Plus). Observe que el primer argumento de estos comandos es el valor $y_0 - r$, no el valor y_0. Esto es porque algunos autores prefieren definir que la distribución binomial negativa sea la de la variable aleatoria Y^* = el número de fracasos *antes del r-ésimo éxito*. En nuestra formulación, la variable aleatoria binomial negativa, Y, se interpreta como el número del intento *en el que ocurre el r-ésimo éxito*. En el Ejercicio 3.100 verá que $Y^* = Y - r$. Debido a esta relación entre las versiones de variables aleatorias binomiales negativas, $P(Y = y_0) = P(Y - r = y_0 - r) = P(Y^\star = y_0 - r)$. R calcula probabilidades asociadas con Y^*, lo que explica por qué los argumentos para `dnbinom` y `pnbinom` son $y_0 - r$ en lugar de y_0.

La media y la varianza de una variable aleatoria con distribución binomial negativa se pueden deducir directamente de las Definiciones 3.4 y 3.5 mediante el uso de técnicas como las ilustradas previamente, pero sumar la serie infinita resultante es tedioso. Estas deducciones serán mucho más fáciles después que hayamos desarrollado algunas de las técnicas del Capítulo 5. Por ahora, expresamos el siguiente teorema sin prueba.

TEOREMA 3.9 Si Y es una variable aleatoria con una distribución binomial negativa,

$$\mu = E(Y) = \frac{r}{p} \quad \text{y} \quad \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

EJEMPLO 3.15 Una gran acumulación de bombas usadas contiene 20% que necesitan ser reparadas. Una trabajadora de mantenimiento es enviada a esas bombas con tres juegos de piezas de reparación. Ella selecciona bombas al azar y las prueba una por una. Si la bomba funciona, la separa para usarla más adelante pero, si no funciona, utiliza uno de los conjuntos de reparación en la bomba. Suponga que le lleva 10 minutos probar una bomba que está en buenas condiciones y 30 minutos para probar y reparar una bomba que no funciona. Encuentre la media y la varianza del tiempo total que tarda la trabajadora para usar sus tres equipos de reparación.

Solución Denote con Y el número del intento en el que se localiza la tercera bomba que no funciona. Se deduce que Y tiene una distribución binomial negativa con $p = .2$. Entonces, $E(Y) = 3/(.2) = 15$ y $V(Y) = 3(.8)/(.2)^2 = 60$. Debido a que se requieren otros 20 minutos para reparar cada bomba defectuosa, el tiempo total necesario para usar los tres equipos de reparación es

$$T = 10Y + 3(20).$$

Usando el resultado obtenido en el Ejercicio 3.33, vemos que

$$E(T) = 10E(Y) + 60 = 10(15) + 60 = 210$$

y

$$V(T) = 10^2 V(Y) = 100(60) = 6000.$$

En consecuencia, el tiempo total necesario para usar los tres equipos de reparación tiene media de 210 y desviación estándar $\sqrt{6000} = 77.46$. ■

Ejercicios

3.90 Los empleados de una empresa que manufactura aislamientos están siendo examinados en busca de indicios de asbesto en sus pulmones. La empresa ha sido requerida para enviar tres empleados que tengan indicios positivos de asbesto a un centro médico para realizarles exámenes adicionales. Si 40% de los empleados tienen indicios positivos de asbesto en sus pulmones, encuentre la probabilidad de que diez empleados deban ser examinados para hallar tres positivos.

3.91 Consulte el Ejercicio 3.90. Si cada examen cuesta \$20, encuentre el valor y la varianza esperados del costo total de realizar los exámenes necesarios para hallar los tres positivos.

3.92 Diez por ciento de los motores fabricados en una línea de ensamble son defectuosos. Si los motores se seleccionan al azar uno a la vez y se prueban, ¿cuál es la probabilidad de que el primer motor no defectuoso sea hallado en el segundo intento?

3.93 Consulte el Ejercicio 3.92. ¿Cuál es la probabilidad de que el tercer motor no defectuoso sea hallado

a en el quinto intento?,

b en el quinto intento o antes?

3.94 Consulte el Ejercicio 3.92. Encuentre la media y la varianza del número del intento en el que

a sea hallado el primer motor no defectuoso,

b sea hallado el tercer motor no defectuoso.

3.95 Consulte el Ejercicio 3.92. Dado que los primeros dos motores probados resultaron defectuosos, ¿cuál es la probabilidad de que al menos dos motores más deban ser probados antes de hallar el primero no defectuoso?

3.96 Las líneas telefónicas que dan servicio a la oficina de reservaciones de una aerolínea están todas ocupadas alrededor de 60% del tiempo.

a Si una persona llama a esta oficina, ¿cuál es la probabilidad de que complete su llamada en el primer intento? ¿En el segundo intento? ¿En el tercero?

b Si usted y un amigo deben ambos completar llamadas a esta oficina, ¿cuál es la probabilidad de que un total de cuatro intentos sean necesarios para que los dos terminen su comunicación?

3.97 Un estudio geológico indica que un pozo petrolero de exploración debe descubrir petróleo con probabilidad .2.

a ¿Cuál es la probabilidad de que el primer hallazgo sea en el tercer pozo perforado?

b ¿Cuál es la probabilidad de que el tercer hallazgo sea en el séptimo pozo perforado?

c ¿Qué suposiciones se hicieron para obtener las respuestas a los incisos a y b?

d Encuentre la media y la varianza del número de pozos que deben ser perforados si la compañía desea abrir tres pozos productores.

***3.98** Considere la distribución binomial negativa dada en la Definición 3.9.

a Demuestre que si $y \geq r+1$, $\dfrac{p(y)}{p(y-1)} = \left(\dfrac{y-1}{y-r}\right)q$. Esto establece una relación repetitiva entre probabilidades binomiales negativas sucesivas, porque $p(y) = p(y-1) \times \left(\dfrac{y-1}{y-r}\right)q$.

b Demuestre que $\dfrac{p(y)}{p(y-1)} = \left(\dfrac{y-1}{y-r}\right)q > 1$ si y $< \dfrac{r-q}{1-q}$. Del mismo modo, $\dfrac{p(y)}{p(y-1)} < 1$ si $y > \dfrac{r-q}{1-q}$.

c Aplique el resultado del inciso b para el caso $r = 7$, $p = .5$ para determinar los valores de y para los que $p(y) > p(y-1)$.

***3.99** En una secuencia de intentos idénticos e independientes con dos posibles resultados en cada intento, S y F, y con $P(S) = p$, ¿cuál es la probabilidad de que exactamente y intentos ocurran *antes* que se presente el r– ésimo éxito?

***3.100** Si Y es una variable aleatoria binomial negativa, defina $Y^* = Y - r$. Si Y se interpreta como el número de intento en el que ocurre el r-ésimo éxito, entonces Y^* se puede interpretar como el número de fracasos *antes* del r-ésimo éxito.

a Si $Y^* = Y - r$, $P(Y^* = y) = P(Y - r = y) = P(Y = y + r)$ para $y = 0, 1, 2, \ldots$ demuestre que $P(Y^* = y) = \dbinom{y+r-1}{r-1} p^r q^y \quad y = 0, 1, 2, \ldots$

b Deduzca la media y la varianza de la variable aleatoria Y^* usando la relación $Y^* = Y - r$, donde Y es binomial negativa y el resultado del Ejercicio 3.33.

***3.101** **a** Observamos una sucesión de intentos idénticos e independientes con dos posibles resultados en cada intento, S y F, y con $P(S) = p$. El número del intento en el que observamos el quinto éxito, Y, tiene una distribución binomial negativa con parámetros $r = 5$ y p. Suponga que observamos el quinto éxito en el onceavo intento. Encuentre el valor de p que *maximice* $P(Y = 11)$.

b Generalice el resultado del inciso a para hallar el valor de p que *maximice* $P(Y = y_0)$ cuando Y tiene una distribución binomial negativa con parámetros r (conocidos) y p.

3.7 La distribución de probabilidad hipergeométrica

En el Ejemplo 3.6 consideramos una población de votantes, 40% de los cuales estaban a favor del candidato Jones. Se seleccionó una muestra de votantes y Y (el número a favor de Jones) había de ser observado. Concluimos que si el tamaño muestral n era pequeño con respecto al tamaño poblacional N, la distribución de Y podría ser calculada por medio de una distribución binomial. También determinamos que si n era grande con respecto a N, la probabilidad *condicional* de seleccionar un votante a favor de Jones en un muestreo posterior sería afectado de manera importante por las preferencias observadas de personas seleccionadas en muestreos anteriores. Entonces los intentos no eran independientes y la distribución de probabilidad para Y no podría ser calculada en forma adecuada por una distribución de probabilidad binomial. Entonces, necesitamos desarrollar la distribución de probabilidad para Y cuando n es grande con respecto a N.

Suponga que una población contiene un número finito N de elementos que posee una de dos características. Así, r de los elementos podrían ser rojos y $b = N - r$, negros. Una muestra de n elementos se selecciona al azar de la población y la variable aleatoria de interés es Y, el número de elementos rojos de la muestra. Esta variable aleatoria tiene lo que se conoce como la *distribución de probabilidad hipergeométrica.* Por ejemplo, el número de trabajadores que son mujeres, Y, en el Ejemplo 3.1 tiene la distribución hipergeométrica.

La distribución de probabilidad hipergeométrica se puede obtener usando los teoremas combinatorios dados en la Sección 2.6 y el método de punto muestral. Un punto muestral del espacio muestral S corresponderá a una selección única de n elementos, algunos rojos y el resto negros. Al igual que en el experimento binomial, cada punto muestral puede ser caracterizado por un n arreglo cuyos elementos correspondan a una selección de n elementos del total de N. Si cada elemento de la población fuera a ser numerado de 1 a N, el punto muestral que indique la selección de artículos 5, 7, 8, 64, 17,..., 87 aparecería como n arreglo.

$$(\underbrace{5, 7, 8, 64, 17, \ldots, 87}_{n \text{ posiciones}}).$$

El número total de puntos muestrales en S, por tanto, será igual al número de formas de seleccionar un subconjunto de n elementos de entre una población de N, o sea $\binom{N}{n}$. Como la selección aleatoria implica que todos los puntos muestrales sean igualmente probables, la probabilidad de un punto muestral en S es

$$P(E_i) = \frac{1}{\binom{N}{n}}, \qquad \text{toda } E_i \in S.$$

El número total de puntos muestrales del evento numérico $Y = y$ es el número de puntos muestrales en S que contenga y elementos rojos y $(n - y)$ negros. Este número se puede obtener al aplicar la regla mn (Sección 2.6). El número de formas de seleccionar y elementos rojos, para llenar y posiciones del n arreglo que representa un punto muestral, es el número de formas de seleccionar y de un total de r, o sea $\binom{r}{y}$. [Usamos la convención $\binom{a}{b}=0$ si $b > a$.] El número total de formas de seleccionar $(n - y)$ elementos negros para llenar las restantes $(n - y)$ posiciones en el n arreglo es el número de formas de seleccionar $(n - y)$ elementos negros de un posible $(N - r)$ o $\binom{N-r}{n-y}$. Entonces el número de puntos muestrales del evento numérico $Y = y$ es el número de formas de combinar un conjunto de y elementos rojos y $(n - y)$ negros. Por la regla mn, este es el producto $\binom{r}{y} \times \binom{N-r}{n-y}$. Sumando las probabilidades de los puntos muestrales del evento numérico $Y = y$ (multiplicando el número de puntos muestrales por la probabilidad común por punto muestral), obtenemos la función de probabilidad hipergeométrica.

DEFINICIÓN **3.10**

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad hipergeométrica* si y sólo si

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

donde y es un entero $0, 1, 2, \ldots, n$, sujeto a las restricciones $y \le r$ y $n - y \le N - r$.

Con la convención $\binom{a}{b} = 0$ si $b > a$, es evidente que $p(y) \ge 0$ para las probabilidades hipergeométricas. El hecho de que las probabilidades hipergeométricas ascienden a 1 se deduce del hecho que

$$\sum_{i=0}^{n} \binom{r}{i}\binom{N-r}{n-i} = \binom{N}{n}.$$

Un bosquejo de la prueba de este resultado aparece en el Ejercicio 3.216.

EJEMPLO **3.16** Un problema importante encontrado por directores de personal y otros que se enfrentan a la selección del mejor candidato en un conjunto finito de elementos, queda ejemplificado en la siguiente situación. De un grupo de 20 ingenieros con título de Ph.D, 10 de ellos son seleccionados al azar para un empleo. ¿Cuál es la probabilidad de que los 10 seleccionados incluyan los cinco mejores ingenieros del grupo de 20?

Solución Para este ejemplo, $N = 20$, $n = 10$ y $r = 5$. Esto es, hay sólo 5 en el conjunto de 5 mejores ingenieros y buscamos la probabilidad de que $Y = 5$, donde Y denota el número de mejores ingenieros entre los diez seleccionados. Entonces

$$p(5) = \frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}} = \left(\frac{15!}{5!10!}\right)\left(\frac{10!10!}{20!}\right) = \frac{21}{1292} = .0162 \qquad \blacksquare$$

Suponga que una población de tamaño N está formada por r unidades con el atributo y $N - r$ sin éste. Si se toma una muestra de tamaño n, sin restitución, y Y es el número de elementos con el atributo en la muestra, $P(Y = y_0) = p(y_0)$ se puede hallar con el comando `dhyper(y0,r,N-r,n)` de *R* (o *S*-Plus). El comando `dhyper(5,5,15,10)` da el valor para $p(5)$ en el Ejemplo 3.16. De manera alternativa, $P(Y \leq y_0)$ se encuentra usando el comando `phyper(y0,r,N-r,n)` de *R* (o *S*-Plus).

La media y la varianza de una variable aleatoria con distribución hipergeométrica se puede deducir directamente de las Definiciones 3.4 y 3.5. No obstante, deducir expresiones de forma cerrada para las sumatorias resultantes es un tanto tedioso. En el Capítulo 5 desarrollaremos métodos que permiten una deducción de los resultados mucho más sencilla que se presenta en el siguiente teorema.

TEOREMA **3.10**

Si Y es una variable aleatoria con distribución hipergeométrica,

$$\mu = E(Y) = \frac{nr}{N} \quad \text{y} \quad \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Aun cuando la media y la varianza de la variable aleatoria hipergeométrica parecen ser más bien complicadas, presentan un sorprendente parecido con la media y la varianza de una variable aleatoria binomial. De hecho, si definimos $p = \frac{r}{N}$ y $q = 1 - p = \frac{N-r}{N}$, podemos expresar también la media y la varianza de la hipergeométrica como $\mu = np$ y

$$\sigma^2 = npq\left(\frac{N-n}{N-1}\right).$$

Se puede ver el factor

$$\frac{N-n}{N-1}$$

en $V(Y)$ como un ajuste que es apropiado cuando n es grande con respecto a N. Para n fija, cuando $N \to \infty$,

$$\frac{N-n}{N-1} \to 1.$$

EJEMPLO **3.17** Un producto industrial se envía en lotes de 20. Es costoso realizar pruebas para determinar si un artículo es defectuoso y, por tanto, el fabricante muestrea su producción en lugar de usar un plan de inspección al 100%. Un plan de muestreo, construido para minimizar el número de piezas defectuosas enviadas a los clientes, exige muestrear cinco artículos de cada lote y rechazar el lote si se observa más de una pieza defectuosa. (Si el lote es rechazado, cada artículo del mismo se prueba posteriormente.) Si un lote contiene cuatro piezas defectuosas, ¿cuál es la probabilidad de que sea rechazado? ¿Cuál es el número esperado de piezas defectuosas en la muestra de tamaño 5? ¿Cuál es la varianza del número de piezas defectuosas de la muestra de tamaño 5?

Solución Sea E igual al número de piezas defectuosas de la muestra. Entonces $N = 20$, $r = 4$ y $n = 5$. El lote será rechazado si $Y = 2$, 3 o 4. Entonces

$$\begin{aligned} P(\text{rechazar el lote}) &= P(Y \geq 2) = p(2) + p(3) + p(4) \\ &= 1 - p(0) - p(1) \\ &= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}} \\ &= 1 - .2817 - .4696 = .2487. \end{aligned}$$

La media y varianza del número de piezas defectuosas de la muestra de tamaño 5 son

$$\mu = \frac{(5)(4)}{20} = 1 \quad \text{y} \quad \sigma^2 = 5\left(\frac{4}{20}\right)\left(\frac{20-4}{20}\right)\left(\frac{20-5}{20-1}\right) = .632. \qquad \blacksquare$$

El Ejemplo 3.17 comprende muestrear un lote de N productos industriales, de los cuales r son defectuosos. La variable aleatoria de interés es Y, el número de defectuosos en la muestra de tamaño n. Como se observó al principio de esta sección, Y posee una distribución aproximadamente binomial cuando N es grande y n es relativamente pequeña. En consecuencia, esperaríamos que las probabilidades asignadas a valores de Y por la distribución hipergeométrica se aproximen a las asignadas por la distribución binomial cuando N se hace grande y r/N, la fracción de piezas defectuosas de la población, se mantiene constante e igual a p. El estudiante puede verificar esta expectativa si usa teoremas de límite, que encontrará en sus cursos de cálculo, para demostrar que

$$\lim_{N \to \infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y (1-p)^{n-y},$$

donde

$$\frac{r}{N} = p.$$

(La prueba de este resultado se omite.) Por tanto, para una fracción fija defectuosa $p = r/N$, la función de probabilidad hipergeométrica converge hacia la función de probabilidad binomial cuando N se hace grande.

Ejercicios

3.102 Una urna contiene diez canicas, de las cuales cinco son verdes, dos son azules y tres son rojas. Tres canicas se van a sacar de la urna, una a la vez sin reemplazo. ¿Cuál es la probabilidad de que las tres canicas sacadas sean verdes?

3.103 Un almacén contiene diez máquinas impresoras, cuatro de las cuales son defectuosas. Una compañía selecciona cinco de las máquinas al azar pensando que todas están en buenas condiciones. ¿Cuál es la probabilidad de que las cinco no sean defectuosas?

3.104 Se tienen 20 paquetes de polvo blanco, de idéntico aspecto, 15 de ellos contienen cocaína y 5 no la tienen. Cuatro paquetes se seleccionaron al azar, se probó su contenido y se encontró que contenían cocaína. Se seleccionaron otros dos paquetes del resto y fueron vendidos por policías secretos a un solo comprador. ¿Cuál es la probabilidad de que los 6 paquetes escogidos al azar sea tal que los primeros 4 contengan cocaína y los 2 vendidos al comprador no la contengan?

3.105 En el sur de California, un número creciente de personas que tratan de obtener credenciales de profesores están escogiendo internados pagados en vez de los programas tradicionales de enseñanza a estudiantes. Un grupo de ocho candidatos para tres plazas de enseñanza locales estaba formado por cinco que se habían inscrito en internados pagados y tres que se inscribieron en programas tradicionales de enseñanza a estudiantes. Los ocho candidatos parecen estar igualmente capacitados, de modo que se seleccionan tres al azar para ocupar las plazas abiertas. Sea Y el número de candidatos capacitados en internados que son contratados.

a ¿Tiene Y una distribución hipergeométrica o binomial? ¿Por qué?

b Encuentre la probabilidad de que sean contratados dos o más candidatos capacitados en internados.

c ¿Cuáles son la media y la desviación estándar de Y?

3.106 Consulte el Ejercicio 3.103. La compañía repara las impresoras defectuosas a un costo de \$50 cada una. Encuentre la media y la varianza del costo total de reparación.

3.107 Un grupo de seis paquetes de software que hay para resolver un problema de programación ha sido clasificado del 1 al 6 (del mejor al peor). Una firma de ingeniería, no informada de la clasificación, selecciona al azar y luego compra dos de los paquetes. Denote con Y el número de paquetes comprados por la empresa que están clasificados 3, 4, 5 o 6. Dé la distribución de probabilidad para Y.

3.108 Un embarque de 20 cámaras incluye 3 que son defectuosas. ¿Cuál es el número mínimo de cámaras que debe ser seleccionado si requerimos que $P(\text{al menos 1 defectuosa}) \geq .8$?

3.109 Es frecuente que las semillas sean tratadas con fungicidas para protegerlas en ambientes húmedos y con desecación defectuosa. Un intento a pequeña escala, que comprende cinco semillas tratadas y cinco no tratadas, fue realizado antes de un experimento a gran escala para explorar cuánto fungicida aplicar. Las semillas se plantaron en un suelo húmedo y se contó el número de plantas que brotaron. Si la solución no era efectiva y cuatro plantas brotaron en realidad, ¿cuál es la probabilidad de que

a las cuatro plantas brotaran de semillas tratadas?

b tres o menos brotaran de semillas tratadas?

c al menos una brotara de semillas no tratadas?

3.110 Una corporación está muestreando sin reemplazo para $n = 3$ firmas para determinar aquella de la cual comprar ciertos abastecimientos. La muestra se ha de seleccionar de un grupo de seis firmas, de las cuales cuatro son locales y dos no. Denote con Y el número de firmas no locales de entre las tres seleccionadas.

a $P(Y = 1)$.

b $P(Y \geq 1)$.

c $P(Y \leq 1)$.

3.111 Las especificaciones exigen que un termistor se pruebe entre 9000 y 10,000 ohms a 25° Celsius. Se dispone de diez termistores y tres de éstos han de ser seleccionados para usarlos. Denote con Y el número de entre los tres que no se apegan a las especificaciones. Encuentre las distribuciones de probabilidad para Y (en forma tabular) dadas las siguientes condiciones:

a Dos termistores no se apegan a las especificaciones de entre los diez disponibles.

b Cuatro termistores no se apegan a las especificaciones de entre los diez disponibles.

3.112 Unas fotocopiadoras usadas son enviadas al proveedor, limpiadas y luego devueltas según convenios de renta. No se hacen reparaciones importantes y en consecuencia algunos clientes reciben máquinas que no funcionan bien. Entre ocho fotocopiadoras disponibles hoy, tres de ellas están funcionando mal. Un cliente desea rentar cuatro máquinas de inmediato. Para satisfacer la fecha límite fijada por el cliente, cuatro de las ocho máquinas se seleccionan al azar y, sin más pruebas, son enviadas al cliente. ¿Cuál es la probabilidad de que el cliente reciba

a máquinas que no funcionen mal?

b al menos una máquina que funcione mal?

3.113 Un jurado de 6 personas fue seleccionado de entre un grupo de 20 miembros de jurado potenciales, de los cuales 8 eran afroamericanos y 12 de raza blanca. Supuestamente, el jurado se seleccionó al azar pero contenía sólo un afroamericano. ¿Piensa el lector que hay alguna razón para dudar de la aleatoriedad de la selección?

3.114 Consulte el Ejercicio 3.113. Si el proceso de selección fuera realmente aleatorio, ¿cuáles serían la media y la varianza del número de afroamericanos seleccionados para el jurado?

3.115 Suponga que un radio contiene seis transistores, dos de los cuales están defectuosos. Se seleccionan al azar tres transistores, se retiran del radio y se inspeccionan. Sea Y igual al número de defectuosos observado, donde $Y = 0$, 1 o 2. Encuentre la distribución de probabilidad para Y. Exprese sus resultados gráficamente como histograma de probabilidad.

3.116 Simule el experimento descrito en el Ejercicio 3.115, al marcar seis canicas o monedas de modo que dos representen defectuosas y cuatro no defectuosas. Ponga las canicas en un sombrero, revuélvalas, saque tres y registre Y, el número de defectuosas observado. Restituya las canicas y repita el proceso hasta que se hayan registrado $n = 100$ observaciones de Y. Construya un histograma de frecuencia relativa para esta muestra y compárelo con la distribución de probabilidad poblacional (Ejercicio 3.115).

3.117 En una producción de línea de ensamble de robots industriales se pueden instalar conjuntos de cajas de engranes en un minuto cada una si los agujeros han sido taladrados correctamente en las cajas y en diez minutos si deben taladrarse agujeros. Hay veinte cajas de engranes en existencia, 2 con agujeros taladrados de manera incorrecta. Cinco cajas de engranes deben seleccionarse de entre las 20 disponibles para instalarse en los siguientes cinco robots.

a Encuentre la probabilidad de que las 5 cajas de engranes ajusten correctamente.

b Encuentre la media, la varianza y la desviación estándar del tiempo que toma instalar estas 5 cajas de engranes.

3.118 Se reparten cinco cartas al azar y sin reemplazo de una baraja de 52 cartas. ¿Cuál es la probabilidad de que la mano contenga los 4 ases si se sabe que contiene al menos 3 ases?

3.119 Se reparten cartas al azar y sin reemplazo de una baraja de 52 cartas. ¿Cuál es la probabilidad de que el segundo rey se reparta en la quinta carta?

***3.120** Los tamaños de poblaciones de animales se calculan en ocasiones con el método de capturar, marcar y recapturar. En este método se capturan k animales, se marcan y luego se sueltan en la población. Cierto tiempo después se capturan n animales y se observa Y, el número de animales marcados de entre los n. Las probabilidades asociadas con Y son una función de N, el número de animales de la población, de modo que el valor observado de Y contiene información sobre esta N desconocida. Suponga que $k = 4$ animales son marcados y luego soltados. Una muestra de $n = 3$ animales se selecciona entonces al azar de entre la misma población. Encuentre $P(Y = 1)$ como función de N. ¿Qué valor de N *maximizará* $P(Y = 1)$?

3.8 La distribución de probabilidad de Poisson

Suponga que deseamos hallar la distribución de probabilidad del número de accidentes automovilísticos ocurridos en un crucero particular durante un periodo de una semana. A primera vista esta variable aleatoria, el número de accidentes, no parece estar ni remotamente relacionada con una variable aleatoria binomial, pero veremos que existe una relación interesante.

Considere el periodo, una semana en este ejemplo, como dividido entre n subintervalos, *cada uno de los cuales es tan pequeño que a lo sumo un accidente podría ocurrir en él con probabilidad diferente de cero.* Denotando con p la probabilidad de un accidente en cualquier subintervalo, tenemos, para todos los fines prácticos,

$$P(\text{no ocurren accidentes en un subintervalo}) = 1 - p,$$

$$P(\text{ocurre un accidente en un subintervalo}) = p,$$

$$P(\text{ocurre más de un accidente en un subintervalo}) = 0.$$

Entonces el número total de accidentes en la semana es precisamente el número total de subintervalos que contienen un accidente. Si la ocurrencia de accidentes puede ser considerada como independiente de un intervalo a otro, el número total de accidentes tiene una distribución binomial.

Aun cuando no hay una forma única de seleccionar los subintervalos y por tanto no conocemos ni n ni p, parece razonable que cuando dividimos la semana en un número mayor de n subintervalos, disminuye la probabilidad p de un accidente en uno de estos subintervalos más cortos. Haciendo $\lambda = np$ y tomando el límite de la probabilidad binomial $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$ cuando $n \to \infty$, tenemos

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{y} p^y (1-p)^{n-y} &= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-y+1)}{y!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} \\
&= \lim_{n\to\infty} \frac{\lambda^y}{y!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\ldots(n-y+1)}{n^y} \left(1 - \frac{\lambda}{n}\right)^{-y} \\
&= \frac{\lambda^y}{y!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y} \left(1 - \frac{1}{n}\right) \\
&\quad \times \left(1 - \frac{2}{n}\right) \times \cdots \times \left(1 - \frac{y-1}{n}\right).
\end{aligned}$$

Si observamos que

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

y todos los otros términos a la derecha del límite tienen un límite de 1, obtenemos

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}.$$

(*Nota:* $e = 2.718\ldots$.) Se dice que las variables aleatorias que poseen esta distribución tienen una distribución de Poisson. En consecuencia, Y, el número de accidentes por semana, tiene la distribución de Poisson que acabamos de deducir.

Debido a que la función de probabilidad binomial converge a la de Poisson, las probabilidades de Poisson se pueden usar para calcular sus similares binomiales para n grande, p pequeña y $\lambda = np$ menor que, aproximadamente, 7. El Ejercicio 3.134 pide al estudiante calcular probabilidades binomiales y de Poisson correspondientes y demostrará lo adecuado del cálculo.

Es frecuente que la distribución de probabilidad de Poisson brinde un buen modelo para la distribución de probabilidad del número Y de eventos raros que ocurren en el espacio, tiempo, volumen o cualquier otra dimensión, donde λ es el valor promedio de Y. Como hemos observado, proporciona un buen modelo para la distribución de probabilidad del número Y de accidentes automovilísticos, industriales y otros tipos en una unidad de tiempo determinada. Otros ejemplos de variables aleatorias con distribuciones aproximadas de Poisson son el número de llamadas telefónicas manejadas por un conmutador en un intervalo, el número de partículas radiactivas que se desintegran en un periodo particular, el número de errores que comete una mecanógrafa al escribir una página y el número de automóviles que usan una rampa de acceso a una autopista en un intervalo de diez minutos.

DEFINICIÓN 3.11 Se dice que una variable aleatoria Y tiene *distribución de probabilidad de Poisson* si y sólo si

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda}, \qquad y = 0, 1, 2, \ldots \quad \lambda > 0.$$

Como veremos en el Teorema 3.11, el parámetro λ que aparece en la fórmula para la distribución de Poisson es en realidad la media de la distribución.

EJEMPLO 3.18 Demuestre que las probabilidades asignadas por la distribución de probabilidad de Poisson satisfacen los requisitos de que $0 \le p(y) \le 1$ para toda y y $\sum_y p(y) = 1$.

Solución Como $\lambda > 0$, es obvio que $p(y) > 0$ para $y = 0, 1, 2, \ldots$, y que $p(y) = 0$ de otro modo. Además,

$$\sum_{y=0}^{\infty} p(y) = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}e^{-\lambda} = e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} = e^{-\lambda}e^{\lambda} = 1$$

porque la suma infinita $\sum_{y=0}^{\infty} \lambda^y/y!$ es una expansión de serie de e^{λ}. En el Apéndice A1.11 se dan sumas de series especiales. ■

EJEMPLO 3.19 Suponga que se diseña un sistema aleatorio de patrulla de policía para que un oficial de patrulla pueda estar en un lugar de su ruta $Y = 0, 1, 2, 3, \ldots$ veces por periodo de media hora, con cada lugar visitado un promedio de una vez por periodo. Suponga que Y posee, aproximadamente, una distribución de probabilidad de Poisson. Calcule la probabilidad de que el oficial de patrulla no llegue a un lugar determinado durante un periodo de media hora. ¿Cuál es la probabilidad de que el lugar sea visitado una vez? ¿Dos veces? ¿Al menos una vez?

Solución Para este ejemplo el periodo es media hora y el número medio de visitas por intervalo de media hora es $\lambda = 1$. Entonces

$$p(y) = \frac{(1)^y e^{-1}}{y!} = \frac{e^{-1}}{y!} \qquad y = 0,\ 1,\ 2, \ldots$$

El evento de que un lugar determinado no sea visitado en un periodo de media hora corresponde a $(Y = 0)$, y

$$P(Y = 0) = p(0) = \frac{e^{-1}}{0!} = e^{-1} = .368.$$

Del mismo modo,

$$p(1) = \frac{e^{-1}}{1!} = e^{-1} = .368,$$

y

$$p(2) = \frac{e^{-1}}{2!} = \frac{e^{-1}}{2} = .184.$$

La probabilidad de que el lugar sea visitado *al menos* una vez es el evento $(Y \geq 1)$. Entonces

$$P(Y \geq 1) = \sum_{y=1}^{\infty} p(y) = 1 - p(0) = 1 - e^{-1} = .632. \qquad \blacksquare$$

Si Y tiene una distribución de Poisson con media λ, $P(Y = y_0) = p(y_0)$ se pueden hallar con el uso del comando `dpois (y0, λ).` de *R* (o s-Plus). Si deseáramos usar *R* para obtener $p(2)$ en el Ejemplo 3.19, usamos el comando `dpois (2, 1)`. Alternativamente, $P(Y \leq y_0)$ se encuentra con el uso del comando de `ppois (y0, λ)` *R* (o *S*-Plus).

EJEMPLO **3.20** Cierto tipo de árbol tiene plantas que han crecido de semillas dispersas al azar en una superficie grande, con la densidad media de plantas siendo aproximadamente de cinco por yarda cuadrada. Si esa zona un guardabosques localiza al azar diez regiones de muestreo de 1 yarda cuadrada, encuentre la probabilidad de que ninguna de las regiones contenga plantas que hayan crecido de semillas.

Solución Si las plantas realmente están dispersas al azar, el número de plantas por región, Y, se puede modelar como una variable aleatoria de Poisson con $\lambda = 5$. (La densidad promedio es de cinco por yarda cuadrada.) Entonces,

$$P(Y = 0) = p(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-5} = .006738.$$

La probabilidad de que $Y = 0$ en diez regiones seleccionadas de manera independiente es $(e^{-5})^{10}$ porque la probabilidad de la intersección de eventos independientes es igual al producto de las probabilidades respectivas. La probabilidad resultante es en extremo pequeña. Entonces, si este evento ocurriera en realidad, cuestionaría seriamente la suposición de aleatoriedad, la densidad promedio de plantas expresada o ambos. $\blacksquare$

Para comodidad del estudiante, damos en la Tabla 3, Apéndice 3, las sumas parciales $\sum_{y=0}^{a} p(y)$ para la distribución de probabilidad de Poisson para muchos valores de λ entre .02 y 25. Esta tabla se ha elaborado de manera similar a la tabla de sumas parciales para la distribución binomial, Tabla 1, Apéndice 3. El siguiente ejemplo ilustra el uso de la Tabla 3 y demuestra que la distribución de probabilidad de Poisson puede aproximar la distribución de probabilidad binomial.

EJEMPLO **3.21** Suponga que Y posee una distribución binomial con $n = 20$ y $p = .1$. Encuentre el valor exacto de $P(Y \le 3)$ usando la tabla de probabilidades binomiales, Tabla 1, Apéndice 3. Use la Tabla 3, Apéndice 3, para aproximar esta probabilidad, usando una probabilidad correspondiente dada por la distribución de Poisson. Compare los valores exacto y aproximado para $P(Y \le 3)$.

Solución De acuerdo con la Tabla 1, Apéndice 3, el valor exacto (hasta tres lugares decimales) de $P(Y \le 3) = .867$. Si W es una variable aleatoria con distribución de Poisson con $\lambda = np = 20(.1) = 2$, los análisis previos indican que $P(Y \le 3)$ es aproximadamente igual a $P(W \le 3)$. La Tabla 3, Apéndice 3, [o el comando `ppois (3, 2)` de R], da $P(W \le 3) = .857$. Entonces, se puede ver que la aproximación de Poisson es bastante buena, dando un valor que difiere del valor exacto en sólo .01. ■

En nuestra deducción de la media y la varianza de una variable aleatoria con distribución de Poisson, de nuevo usamos la propiedad fundamental que $\sum_y p(y) = 1$ para cualquier distribución de probabilidad discreta.

TEOREMA **3.11** Si Y es una variable aleatoria que posee una distribución de Poisson con parámetro λ, entonces

$$\mu = E(Y) = \lambda \qquad \text{y} \qquad \sigma^2 = V(Y) = \lambda.$$

Demostración Por definición,

$$E(Y) = \sum_y yp(y) = \sum_{y=0}^{\infty} y\,\frac{\lambda^y e^{-\lambda}}{y!}.$$

Observe que el primer término de esta suma es igual a 0 (cuando $y = 0$) y, por tanto,

$$E(Y) = \sum_{y=1}^{\infty} y\,\frac{\lambda^y e^{-\lambda}}{y!} = \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!}.$$

Así como está, esta cantidad no es igual a la suma de los valores de una función de probabilidad $p(y)$ para todos los valores de y, pero podemos cambiarla a la forma apropiada al factorizar λ de la expresión y haciendo $z = y - 1$. Entonces los límites de sumatoria se convierten en $z = 0$ (cuando $y = 1$) y $z = \infty$ (cuando $y = \infty$), y

$$E(Y) = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!}.$$

Observe que $p(z) = \lambda^z e^{-\lambda}/z!$ es la función de probabilidad para una variable aleatoria de Poisson, y $\sum_{z=0}^{\infty} p(z) = 1$. Por tanto, $E(Y) = \lambda$. Entonces, la media de una variable aleatoria de Poisson es el parámetro individual λ que aparece en la expresión para la función de probabilidad de Poisson.

Dejamos la obtención de la varianza como Ejercicio 3.138.

Una forma común de encontrar una variable aleatoria con una distribución Poisson es por medio de un modelo llamado *proceso Poisson,* que es un modelo apropiado para situaciones como la que se describe al principio de esta sección. Si observamos un proceso Poisson y λ es el número medio de sucesos *por unidad* (longitud, área, etc.), entonces Y = número de sucesos en a unidades tiene una distribución Poisson con media $a\lambda$. Una suposición clave en el desarrollo de la teoría del proceso Poisson es la independencia de los números de sucesos en intervalos inconexos (áreas, etc.). Vea en la obra de Hogg, Craigla, y McKean (2005) un desarrollo teórico del proceso Poisson.

EJEMPLO **3.22** Ocurren accidentes industriales de acuerdo con un proceso Poisson con un promedio de tres accidentes por mes. Durante los últimos dos meses ocurrieron diez accidentes. ¿Este número parece altamente improbable si el número medio de accidentes por mes, μ, es todavía igual a 3? ¿Indica un aumento en el número medio de accidentes por mes?

Solución El número de accidentes en *dos* meses, Y, tiene una distribución de probabilidad de Poisson con media $\lambda^* = 2(3) = 6$. La probabilidad de que Y sea de hasta 10 es

$$P(Y \geq 10) = \sum_{y=10}^{\infty} \frac{6^y e^{-6}}{y!}.$$

El tedioso cálculo necesario para hallar $P(Y \geq 10)$ se puede evitar con el uso de la Tabla 3, Apéndice 3, de software como *R* [`ppois (9, 6)` da $P(Y \leq 9)$] o la regla empírica. Del Teorema 3.11,

$$\mu = \lambda^* = 6, \qquad \sigma^2 = \lambda^* = 6, \qquad \sigma = \sqrt{6} = 2.45.$$

La regla empírica nos dice que deberíamos esperar que Y tome valores en el intervalo $\mu \pm 2\sigma$ con una alta probabilidad.

Advierta que $\mu + 2\sigma = 6 + (2)(2.45) = 10.90$. El número observado de accidentes, $Y = 10$, no está a más de 2σ de μ, pero está cerca de la frontera. Por tanto, el resultado observado no es altamente improbable, pero puede ser suficientemente improbable para garantizar una investigación. Vea en el Ejercicio 3.210 la probabilidad exacta $P(|Y - \lambda| \leq 2\sigma)$. ■

Ejercicios

3.121 Denote con Y una variable aleatoria que tenga una distribución de Poisson con media $\lambda = 2$. Encuentre

a $P(Y = 4)$.
b $P(Y \geq 4)$.
c $P(Y < 4)$.
d $P(Y \geq 4\ Y| \geq 2)$.

3.122 Llegan clientes a un mostrador de salida en una tienda de departamentos de acuerdo con una distribución de Poisson, a un promedio de siete por hora. Durante una hora determinada, ¿cuáles son las probabilidades de que

a no lleguen más de tres clientes?,

b lleguen al menos dos clientes?,

c lleguen exactamente cinco clientes?

3.123 La variable aleatoria Y tiene una distribución de Poisson y es tal que $p(0) = p(1)$. ¿Cuál es $p(2)$?

3.124 Aproximadamente 4% de las obleas de silicio producidas por un fabricante tienen menos de dos defectos grandes. Si Y, el número de defectos por oblea, tiene una distribución de Poisson, ¿qué proporción de las obleas tiene más de cinco defectos grandes? [*Sugerencia:* use la Tabla 3, Apéndice 3.]

3.125 Consulte el Ejercicio 3.122. Si se requieren alrededor de diez minutos para servir a cada cliente, encuentre la media y la varianza del tiempo total de servicio para clientes que lleguen durante un periodo de 1 hora. (Suponga que hay un número suficiente de dependientes para que el cliente no tenga que esperar ser atendido.) ¿Es probable que el tiempo total de servicio exceda de 2.5 horas?

3.126 Consulte el Ejercicio 3.122. Suponga que ocurren llegadas de acuerdo con un proceso de Poisson con un promedio de siete por hora. ¿Cuál es la probabilidad de que exactamente dos clientes lleguen en dos horas entre

a las 2:00 p.m. y las 4:00 p.m. (un periodo continuo de dos horas)?,

b la 1:00 p.m. y las 2:00 p.m. o entre las 3:00 p.m. y las 4:00 p.m. (dos periodos de una hora separados que totalizan dos horas)?

3.127 El número de errores mecanográficos hechos por una secretaria tiene una distribución de Poisson con un promedio de cuatro errores por página. Si en una página se dan más de cuatro errores, la secretaria debe volver a escribir toda la página. ¿Cuál es la probabilidad de que una página seleccionada al azar no tenga que volver a ser escrita?

3.128 Llegan autos a una caseta de pago de peaje de acuerdo con un proceso de Poisson con media de 80 autos por hora. Si el empleado hace una llamada telefónica de 1 minuto, ¿cuál es la probabilidad de que al menos 1 auto llegue durante la llamada?

3.129 Consulte el Ejercicio 3.128. ¿Cuánto puede durar la llamada telefónica del empleado si la probabilidad es al menos .4 de que no lleguen autos durante la llamada?

3.130 Un lote de estacionamiento tiene dos entradas. Llegan autos a la entrada I de acuerdo con una distribución de Poisson a un promedio de tres por hora y a la entrada II de acuerdo con una distribución de Poisson a un promedio de cuatro por hora. ¿Cuál es la probabilidad de que un total de tres autos lleguen al lote de estacionamiento en una hora determinada? (Suponga que los números de autos que llega a las dos entradas son independientes.)

3.131 El número de nudos en un tipo particular de madera tiene una distribución de Poisson con un promedio de 1.5 nudos en 10 pies cúbicos de madera. Encuentre la probabilidad de que un bloque de 10 pies cúbicos de madera tenga a lo sumo 1 nudo.

3.132 El número medio de automóviles que entran al túnel de una montaña por periodo de dos minutos es uno. Un número excesivo de autos que entren al túnel durante un breve tiempo produce una situación peligrosa.

Encuentre la probabilidad de que el número de autos que entran durante un periodo de dos minutos exceda de tres. ¿El modelo de Poisson parece razonable para este problema?

3.133 Suponga que el túnel del Ejercicio 3.132 se observa durante diez intervalos de dos minutos, dando así diez observaciones independientes $Y_1, Y_2, \ldots, Y_{10}$, en la variable aleatoria de Poisson. Encuentre la probabilidad de que $Y > 3$ durante al menos uno de los diez intervalos de dos minutos.

3.134 Considere un experimento binomial para $n = 20, p = .05$. Use la Tabla 1, Apéndice 3, para calcular las probabilidades binomiales para $Y = 0, 1, 2, 3$ y 4. Calcule las mismas probabilidades usando la aproximación de Poisson con $\lambda = np$. Compare.

3.135 Un vendedor ha encontrado que la probabilidad de una venta en un solo contacto es aproximadamente .03. Si el vendedor hace contacto con 100 posibles clientes, ¿cuál es la probabilidad aproximada de hacer al menos una venta?

3.136 Más investigación y análisis se han concentrado en el número de enfermedades en las que aparece el organismo *Escherichia coli* (10257:H7), que causa una ruptura de células sanguíneas y hemorragia intestinal en sus víctimas (http://www.hsus.org/ace/11831, marzo 24, de 2004). Esporádicos brotes de *E.coli* han aparecido en Colorado a razón de aproximadamente 2.4 por 100,000 durante un periodo de dos años.

a Si este índice no ha cambiado y si 100,000 casos de Colorado se revisan para este año, ¿cuál es la probabilidad de que se observen al menos 5 casos de *E.coli*?

b Si 100,000 casos de Colorado se revisan para este año y el número de casos de *E.coli* excede de 5, ¿es de esperarse que haya cambiado la media estatal del índice de *E.coli*? Explique.

3.137 La probabilidad de que un ratón inoculado con un suero contraiga cierta enfermedad es .2. Usando la aproximación de Poisson, encuentre la probabilidad de que al menos 3 de entre 30 ratones inoculados contraigan la enfermedad.

3.138 Sea Y que tiene una distribución de Poisson con media λ. Encuentre $E[Y(Y-1)]$ y luego use esto para demostrar que $V(Y) = \lambda$.

3.139 En la producción diaria de cierta clase de cuerda, el número de defectos por pie Y se supone que tiene una distribución de Poisson con media $\lambda = 2$. La utilidad por pie cuando se venda la cuerda está dada por X, donde $X = 50 - 2Y - Y^2$. Encuentre la utilidad esperada por pie.

***3.140** El propietario de una tienda ha abarrotado cierto artículo y decide usar la siguiente promoción para disminuir la oferta. El artículo tiene un precio marcado de \$100. Por cada cliente que compre el artículo durante un día en particular, el propietario reducirá el precio en un factor de un medio. Entonces, el primer cliente pagará \$50 por el artículo, el segundo pagará \$25 y así sucesivamente. Suponga que el número de clientes que compren el artículo durante el día tiene una distribución de Poisson con media 2. Encuentre el costo esperado del artículo al final del día. [*Sugerencia:* el costo al final del día es $100(1/2)^Y$, donde Y es el número de clientes que han comprado el artículo.]

3.141 Un fabricante de alimentos usa una máquina de moldeo por inyección (que produce galletas del tamaño de un bocado y botanas) que proporciona un ingreso para la empresa a razón de \$200 por hora cuando está en operación. No obstante, la máquina se descompone a un promedio de dos veces por cada día que trabaja. Si Y denota el número de descomposturas por día, el ingreso diario generado por la máquina es $R = 1600 - 50Y^2$. Encuentre el ingreso diario esperado por usar la máquina.

***3.142** Denote con $p(y)$ la función de probabilidad asociada con una variable aleatoria de Poisson con media λ.

a Demuestre que la relación entre probabilidades sucesivas satisface la igualdad $\dfrac{p(y)}{p(y-1)} = \dfrac{\lambda}{y}$, para $y = 1, 2, \ldots$

b ¿Para qué valores de y es $p(y) > p(y-1)$?

c Observe que el resultado del inciso a implica que las probabilidades de Poisson aumentan por un tiempo cuando y aumenta y disminuyen de ahí en adelante. Demuestre que $p(y)$ es maximizada cuando $y =$ al máximo entero menor o igual que λ.

3.143 Consulte el Ejercicio 3.142 c. Si el número de llamadas telefónicas al departamento de bomberos, Y, en un día tiene una distribución de Poisson con media de 5.3, ¿cuál es el número más probable de llamadas telefónicas al departamento de bomberos en cualquier día?

3.144 Consulte los ejercicios 3.142 y 3.143. Si el número de llamadas telefónicas al departamento de bomberos, Y, en un día tiene una distribución Poisson con media 6, demuestre que $p(5) = p(6)$ de modo que 5 y 6 son los *dos* valores más probables para Y.

3.9 Momentos y funciones generadoras de momento

Los parámetros μ y σ son medidas descriptivas numéricas significativas que ubican el centro y describen la dispersión asociada con los valores de una variable aleatoria Y, pero no dan una caracterización única de la distribución de Y. Muchas distribuciones diferentes poseen las mismas medias y desviaciones estándar. A continuación consideramos un conjunto de medidas descriptivas numéricas que (al menos en ciertas condiciones) determinan $p(y)$ de manera única.

DEFINICIÓN **3.12** El k-ésimo *momento de una variable aleatoria Y tomada alrededor del origen* se define como $E(Y^k)$ y se denota con μ'_k.

Observe en particular que el primer momento alrededor del origen es $E(Y) = \mu'_1 = \mu$ y que $\mu'_2 = E(Y^2)$ se emplea en el Teorema 3.6 para hallar σ^2.

Otro momento útil de una variable aleatoria es el tomado alrededor de su media.

DEFINICIÓN **3.13** El k-ésimo *momento de una variable aleatoria Y tomado alrededor de su media* o el k-ésimo *momento central de Y*, se define como $E[(Y - \mu)^k]$ y está denotado por μ_k.

En particular, $\sigma^2 = \mu_2$.

Concentremos nuestra atención en los momentos μ'_k alrededor del origen donde $k = 1, 2, 3, \ldots$ Suponga que dos variables aleatorias Y y Z poseen momentos finitos con $\mu'_{1Y} = \mu'_{1Y}, \mu'_{2Y} = \mu'_{2Z}, \ldots, \mu'_{jY} = \mu'_{jZ}$, donde j puede tomar cualquier valor entero. Esto es, las dos variables aleatorias poseen momentos correspondientes idénticos alrededor del origen. En algunas condiciones más bien generales, se puede demostrar que Y y Z tienen distribuciones de probabilidad idénticas. Así, un uso importante de los momentos es para calcular la distribución de probabilidad de una variable aleatoria (por lo general un estimador o tomador de decisiones). Por tanto, los momentos μ'_k., donde $k = 1, 2, 3, \ldots$, son principalmente de valor teórico para $k > 3$.

Otra expectativa interesante es la función generadora de momento para una variable aleatoria, que hablando en forma figurada, compacta todos los momentos para una variable aleatoria en una sola expresión. Definiremos primero la función generadora de momento y luego vamos a explicar la forma en que trabaja.

DEFINICIÓN **3.14**

La *función generadora de momento* $m(t)$ *para una variable aleatoria* Y se define como $m(t) = E(e^{tY})$. Decimos que una función generadora de momento para Y existe si existe una constante positiva b tal que $m(t)$ es finita para $|t| \leq b$.

¿Por qué $E(e^{tY})$ recibe el nombre de *función generadora de momento para* Y? De una expansión de serie para e^{ty}, tenemos

$$e^{ty} = 1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \frac{(ty)^4}{4!} + \cdots.$$

Entonces, suponiendo que μ'_k es finita para $k = 1, 2, 3, \ldots,$ tenemos

$$\begin{aligned} E(e^{tY}) &= \sum_y e^{ty} p(y) = \sum_y \left[1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \cdots\right] p(y) \\ &= \sum_y p(y) + t \sum_y y p(y) + \frac{t^2}{2!} \sum_y y^2 p(y) + \frac{t^3}{3!} \sum_y y^3 p(y) + \cdots \\ &= 1 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \frac{t^3}{3!}\mu'_3 + \cdots. \end{aligned}$$

Este argumento comprende un intercambio de sumatorias, que es justificable si $m(t)$ existe. Entonces, $E(e^{tY})$ es una función de todos los momentos μ'_k alrededor del origen, para $k = 1, 2, 3, \ldots$. En particular, μ'_k es el coeficiente de $t^k/k!$ en la expansión de serie de $m(t)$.

La función generadora de momento posee dos aplicaciones importantes. Primero, si podemos hallar $E(e^{tY})$, podemos hallar cualquiera de los momentos para Y.

TEOREMA **3.12**

Si $m(t)$ existe, entonces para cualquier entero positivo k,

$$\left.\frac{d^k m(t)}{dt^k}\right]_{t=0} = m^{(k)}(0) = \mu'_k.$$

En otras palabras, si el estudiante encuentra la k-ésima derivada de $m(t)$ con respecto a t y luego hace $t = 0$, el resultado será μ'_k.

Demostración

$d^k m(t)/dt^k$, o $m^{(k)}(t)$, es la k-ésima derivada de $m(t)$ con respecto a t. Como

$$m(t) = E(e^{tY}) = 1 + t\mu'_1 + \frac{t^2}{2!}\mu'_2 + \frac{t^3}{3!}\mu'_3 + \cdots,$$

se deduce que

$$\begin{aligned} m^{(1)}(t) &= \mu'_1 + \frac{2t}{2!}\mu'_2 + \frac{3t^2}{3!}\mu'_3 + \cdots, \\ m^{(2)}(t) &= \mu'_2 + \frac{2t}{2!}\mu'_3 + \frac{3t^2}{3!}\mu'_4 + \cdots, \end{aligned}$$

y, en general,

$$m^{(k)}(t) = \mu'_k + \frac{2t}{2!}\mu'_{k+1} + \frac{3t^2}{3!}\mu'_{k+2} + \cdots.$$

Si hacemos $t = 0$ en cada una de las derivadas anteriores, obtenemos

$$m^{(1)}(0) = \mu'_1, \qquad m^{(2)}(0) = \mu'_2,$$

y, en general,

$$m^{(k)}(0) = \mu'_k.$$

Estas operaciones implican intercambiar derivadas y sumas infinitas, que se pueden justificar si existe $m(t)$.

EJEMPLO **3.23** Encuentre la función generadora de momento $m(t)$ para una variable aleatoria con distribución de Poisson y media λ.

Solución

$$m(t) = E(e^{tY}) = \sum_{y=0}^{\infty} e^{ty} p(y) = \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y e^{-\lambda}}{y!}$$

$$= \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y e^{-\lambda}}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!}.$$

Para completar la sumatoria, consulte el Apéndice A1.11 para hallar la expansión de la serie de Taylor

$$\sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} = e^{\lambda e^t}$$

o utilice el método del Teorema 3.11. Así, multiplique y divida por $e^{\lambda e^t}$. Entonces

$$m(t) = e^{-\lambda} e^{\lambda e^t} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y e^{-\lambda e^t}}{y!}.$$

La cantidad a la derecha del signo de sumatoria es la función de probabilidad para una variable aleatoria de Poisson con media λe^t. Así,

$$\sum_{y} p(y) = 1 \quad \text{y} \quad m(t) = e^{-\lambda} e^{\lambda e^t}(1) = e^{\lambda(e^t - 1)}.$$ ■

Los cálculos en el Ejemplo 3.23 no son más difíciles que los del Teorema 3.11, donde sólo se calculó el valor esperado para una variable aleatoria Y de Poisson. La evaluación directa de la varianza de Y por medio del uso del Teorema 3.6 requirió que $E(Y^2)$ se hallara al sumar otra serie [en realidad, obtuvimos $E(Y^2)$ de $E[Y(Y-1)]$ en el Ejercicio 3.138]. El Ejemplo 3.24 ilustra el uso de la función generadora de momento de la variable aleatoria de Poisson para calcular su media y su varianza.

EJEMPLO **3.24** Use la función generadora de momento del Ejemplo 3.23 y el Teorema 3.12 para hallar la media, μ, y la varianza, σ^2, para la variable aleatoria de Poisson.

Solución De acuerdo con el Teorema 3.12, $\mu = \mu'_1 = m^{(1)}(0)$ y $\mu'_2 = m^{(2)}(0)$. Evaluando la primera y segunda derivadas de $m(t)$, obtenemos

$$
\begin{aligned}
m^{(1)}(t) &= \frac{d}{dt}[e^{\lambda(e^t-1)}] = e^{\lambda(e^t-1)} \cdot \lambda e^t, \\
m^{(2)}(t) &= \frac{d^2}{dt^2}[e^{\lambda(e^t-1)}] = \frac{d}{dt}[e^{\lambda(e^t-1)} \cdot \lambda e^t] \\
&= e^{\lambda(e^t-1)} \cdot (\lambda e^t)^2 + e^{\lambda(e^t-1)} \cdot \lambda e^t.
\end{aligned}
$$

Entonces, como

$$
\begin{aligned}
\mu = m^{(1)}(0) &= \left\{ e^{\lambda(e^t-1)} \lambda \; e^t \right\}\Big]_{t=0} = \lambda, \\
\mu'_2 = m^{(2)}(0) &= \left\{ e^{\lambda(e^t-1)} . (\lambda e^t)^2 + e^{\lambda(e^t-1)} . \lambda e^t \right\}\Big]_{t=0} = \lambda^2 + \lambda,
\end{aligned}
$$

el Teorema 3.6 nos dice que $\sigma^2 = E(Y^2) - \mu^2 = \mu'_2 - \mu^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda$. Observe con qué facilidad obtuvimos μ'_2 a partir de $m(t)$. ■

La segunda aplicación (pero primaria) de una función generadora de momento es demostrar que una variable aleatoria posee una distribución de probabilidad particular $p(y)$. Si $m(t)$ existe para una distribución de probabilidad $p(y)$, es única. También, si las funciones generadoras de momento para dos variables aleatorias Y y Z son iguales (para toda $|t| < b$ para alguna $b > 0$), entonces Y y Z deben tener la misma distribución de probabilidad. Se deduce que, si podemos reconocer la función generadora de momento de una variable aleatoria Y como una asociada con una distribución específica, entonces Y debe tener esa distribución.

En resumen, una función generadora de momentos es una expresión matemática que en ocasiones (pero no siempre) proporciona una forma fácil de hallar momentos asociados con variables aleatorias. Lo más importante es que puede usarse para establecer la equivalencia de dos distribuciones de probabilidad.

EJEMPLO **3.25** Suponga que Y es una variable aleatoria con función generadora de momento $m_Y(t) = e^{3.2(e^t-1)}$ ¿Cuál es la distribución de Y?

Solución En el Ejemplo 3.23 demostramos que la función generadora de momento de una variable aleatoria con distribución de Poisson y media λ es $m(t) = e^{\lambda(e^t-1)}$. Observe que la función generadora de momento de Y es exactamente igual a la función generadora de momento de una variable aleatoria con distribución de Poisson y $\lambda = 3.2$. Como las funciones generadoras de momento son únicas, Y *debe* tener una distribución de Poisson con media 3.2. ■

Ejercicios

3.145 Si Y tiene una distribución binomial con n intentos y probabilidad de éxito p, demuestre que la función generadora de momento para Y es

$$m(t) = (pe^t + q)^n, \qquad \text{donde } q = 1 - p.$$

3.146 Derive la función generadora de momento del Ejercicio 3.145 para hallar $E(Y)$ y $E(Y^2)$. Entonces encuentre $V(Y)$.

3.147 Si Y tiene una distribución geométrica con probabilidad de éxito p, demuestre que la función generadora de momento para Y es

$$m(t) = \frac{pe^t}{1 - qe^t}, \qquad \text{donde } q = 1 - p.$$

3.148 Derive la función generadora de momento del Ejercicio 3.147 para hallar $E(Y)$ y $E(Y^2)$. Entonces encuentre $V(Y)$.

3.149 Consulte el Ejercicio 3.145. Use la unicidad de las funciones generadoras de momento para dar la distribución de una variable aleatoria con función generadora de momento $m(t) = (.6e^t + .4)^3$.

3.150 Consulte el Ejercicio 3.147. Use la unicidad de las funciones generadoras de momento para dar la distribución de una variable aleatoria con función generadora de momento $m(t) = \dfrac{.3e^t}{1 - .7e^t}$.

3.151 Consulte el Ejercicio 3.145. Si Y tiene una función generadora de momento $m(t) = (.7e^t + .3)^{10}$, ¿cuál es $P(Y \le 5)$?

3.152 Consulte el Ejercicio 3.23. Si Y tiene una función generadora de momento $m(t) = e^{6(e^t - 1)}$, ¿cuál es $P(|Y - \mu| \le 2\sigma)$?

3.153 Encuentre las distribuciones de las variables aleatorias que tienen cada una de las siguientes funciones generadoras de momento:

a $m(t) = [(1/3)e^t + (2/3)]^5$.

b $m(t) = \dfrac{e^t}{2 - e^t}$.

c $m(t) = e^{2(e^t - 1)}$.

3.154 Consulte el Ejercicio 3.153. Por inspección obtenga la media y la varianza de las variables aleatorias asociadas con las funciones generadoras de momento dadas en los incisos a, b y c.

3.155 Sea $m(t) = (1/6)e^t + (2/6)e^{2t} + (3/6)e^{3t}$. Encuentre lo siguiente:

a $E(Y)$.

b $V(Y)$.

c La distribución de Y.

3.156 Suponga que Y es una variable aleatoria con una función $m(t)$ generadora de momento.

a ¿Cuál es $m(0)$?

b Si $W = 3Y$, demuestre que la función generadora de momento de W es $m(3t)$.

c Si $X = Y - 2$, demuestre que la función generadora de momento de X es $e^{-2t}m(t)$.

3.157 Consulte el Ejercicio 3.156.

a Si $W = 3Y$, use la función generadora de momento de W para demostrar que $E(W) = 3E(Y)$ y $V(W) = 9V(Y)$.

b Si $X = Y - 2$, use la función generadora de momento de X para demostrar que $E(X) = E(Y) - 2$ y $V(X) = V(Y)$.

3.158 Si Y es una variable aleatoria con función generadora de momento $m(t)$ y si W está dada por $W = aY + b$, demuestre que la función generadora de momento de W es $e^{tb}\, m(at)$.

3.159 Use el resultado del Ejercicio 3.158 para demostrar que, si $W = aY + b$, entonces $E(W) = aE(Y) + b$ y $V(W) = a^2V(Y)$.

3.160 Suponga que Y es una variable aleatoria binomial basada en n intentos con probabilidad de éxito p y sea $Y^*(W) = n - Y$.

a Use el resultado del Ejercicio 3.159 para demostrar que $E(Y^*) = nq$ y $V(Y^*) = npq$, donde $q = 1 - p$.

b Use el resultado del Ejercicio 3.158 para demostrar que la función generadora de momento de Y^* es, $m^*(t) = (qe^t + p)^n$ n donde $q = 1 - p$.

c Con base en su respuesta al inciso b, ¿cuál es la distribución de Y^*?

d Si Y se interpreta como el número de éxitos en una muestra de tamaño n, ¿cuál es la interpretación de Y^*?

e Con base en su respuesta al inciso d, ¿por qué las respuestas a los incisos a, b y c son "obvias"?

3.161 Consulte los Ejercicios 3.147 y 3.158. Si Y tiene una distribución geométrica con probabilidad de éxito p, considere $Y^* = Y - 1$. Demuestre que la función generadora de momento de Y^* es$(m^*) = \dfrac{p}{1 - qe}$, donde $q = 1 - p$.

***3.162** Sea $r(t) = \ln[m(t)]$ y denote con $r^{(k)}(0)$ a la k-ésima derivada de $r(t)$ evaluada para $t = 0$. Demuestre que $r^{(1)}(0) = \mu'_1 = \mu$ y $r^{(2)}(0) = \mu'_2 - (\mu'_1)^2 = \sigma^2$. [*Sugerencia:* $m(0) = 1$.]

***3.163** Use los resultados del Ejercicio 3.162 para hallar la media y la varianza de una variable aleatoria de Poisson con $m(t) = e^{5(e^t - 1)}$. Observe que $r(t)$ es más fácil de derivar que $m(t)$ en este caso.

3.10 Funciones generadoras de probabilidad (opcional)

Una clase importante de variables aleatorias discretas es aquella en la que Y representa una cantidad y, en consecuencia, toma valores enteros: $Y = 0, 1, 2, 3, \ldots$ Las variables aleatorias binomiales, geométricas, hipergeométricas y de Poisson caen todas en esta clase. Los siguientes ejemplos dan situaciones prácticas que resultan en variables aleatorias de valores enteros. Uno, que comprende la teoría de colas (líneas de espera), se refiere al número de personas (u objetos) que esperan servicio en un punto particular en el tiempo. El conocimiento del comportamiento de esta variable aleatoria es importante para el diseño de plantas de manufactura en donde la producción consiste en una secuencia de operaciones, cada una de ellas tomando un tiempo diferente para completarse. Un número insuficiente de estaciones de servicio para una operación particular de producción puede resultar en un cuello de botella, la formación de una cola de productos que esperan recibir servicio y en la reducción en la operación de manufactura. La teoría de colas es también importante para determinar el número de cajeras para un supermercado y para diseñar hospitales y clínicas.

Las variables aleatorias de valores enteros también son importantes en estudios de crecimiento de población. Por ejemplo, algunos epidemiólogos están interesados en el crecimiento de poblaciones de bacterias y el crecimiento del número de personas afectadas por una enfermedad en particular. Los números de elementos en cada una de estas poblaciones son variables aleatorias de número entero.

Un procedimiento matemático útil para hallar las distribuciones de probabilidad y otras propiedades de variables aleatorias de valor entero es la función generadora de probabilidad.

DEFINICIÓN 3.15

Sea Y una variable aleatoria de valor entero para la cual $P(Y = i) = p_i$, donde $i = 0, 1, 2, \ldots$. La *función generadora de probabilidad* $P(t)$ para Y se define como

$$P(t) = E(t^Y) = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{i=0}^{\infty} p_i t^i$$

para todos los valores de t tales que $P(t)$ sea finita.

La razón para llamar a $P(t)$ una función generadora de probabilidad es clara cuando comparamos $P(t)$ con la función generadora de momento $m(t)$. En particular, el coeficiente de t^i en $P(t)$ es la probabilidad p_i. De manera correspondiente, el coeficiente de t^i para $m(t)$ es una constante multiplicada por el i-ésimo momento μ_i'. Si conocemos $P(t)$ y podemos expandirlo en una serie, podemos determinar $p(y)$ como el coeficiente de t^y.

La derivación repetida de $P(t)$ da *momentos factoriales* para la variable aleatoria Y.

DEFINICIÓN 3.16

El k-ésimo *momento factorial* para una variable aleatoria Y se define como

$$\mu_{[k]} = E[Y(Y-1)(Y-2)\cdots(Y-k+1)],$$

donde k es un entero positivo.

Nótese que $\mu_{[1]} = E(Y) = \mu$. El segundo momento factorial, $\mu_{[2]} = E[Y(Y-1)]$, fue útil para hallar la varianza para variables aleatorias binomiales, geométricas y de Poisson en el Teorema 3.7, el Ejercicio 3.85 y el Ejercicio 3.138, respectivamente.

TEOREMA 3.13

Si $P(t)$ es la función generadora de probabilidad para una variable aleatoria de valor entero, Y, entonces el k-ésimo momento factorial de Y está dado por

$$\left.\frac{d^k P(t)}{dt^k}\right]_{t=1} = P^{(k)}(1) = \mu_{[k]}.$$

Demostración

Como

$$P(t) = p_0 + p_1 t + p_2 t^2 + p_3 t^3 + p_4 t^4 + \cdots,$$

se deduce que

$$P^{(1)}(t) = \frac{dP(t)}{dt} = p_1 + 2p_2 t + 3p_3 t^2 + 4p_4 t^3 + \cdots,$$

$$P^{(2)}(t) = \frac{d^2 P(t)}{dt^2} = (2)(1)p_2 + (3)(2)p_3 t + (4)(3)p_4 t^2 + \cdots,$$

y, en general,

$$P^{(k)}(t) = \frac{d^k P(t)}{dt^k} = \sum_{y=k}^{\infty} y(y-1)(y-2)\cdots(y-k+1)p(y)t^{y-k}.$$

Haciendo $t = 1$ en cada una de estas derivadas, obtenemos

$$P^{(1)}(1) = p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots = \mu_{[1]} = E(Y),$$
$$P^{(2)}(1) = (2)(1)p_2 + (3)(2)p_3 + (4)(3)p_4 + \cdots = \mu_{[2]} = E[Y(Y-1)],$$

y, en general,

$$\begin{aligned} P^{(k)}(1) &= \sum_{y=k}^{\infty} y(y-1)(y-2)\cdots(y-k+1)p(y) \\ &= E[Y(Y-1)(Y-2)\cdots(Y-k+1)] = \mu_{[k]}. \end{aligned}$$

EJEMPLO **3.26** Encuentre la función generadora de probabilidad para una variable aleatoria geométrica.

Solución Observe que $p_0 = 0$ porque Y no puede tomar este valor. Entonces

$$\begin{aligned} P(t) = E(t^Y) &= \sum_{y=1}^{\infty} t^y q^{y-1} p = \sum_{y=1}^{\infty} \frac{p}{q}(qt)^y \\ &= \frac{p}{q}[qt + (qt)^2 + (qt)^3 + \cdots]. \end{aligned}$$

Los términos de la serie son los de una progresión geométrica infinita. Si $qt < 1$, entonces

$$P(t) = \frac{p}{q}\left(\frac{qt}{1-qt}\right) = \frac{pt}{1-qt}, \qquad \text{si } t < 1/q.$$

(Para la sumatoria de la serie, consulte el Apéndice A1.11.) ■

EJEMPLO **3.27** Use $P(t)$, Ejemplo 3.26, para hallar la media de una variable aleatoria geométrica.

Solución Del Teorema 3.13, $\mu_{[1]} = \mu = P^{(1)}(1)$. Usando el resultado del Ejemplo 3.26,

$$P^{(1)}(t) = \frac{d}{dt}\left(\frac{pt}{1-qt}\right) = \frac{(1-qt)p - (pt)(-q)}{(1-qt)^2}.$$

Haciendo $t = 1$, obtenemos

$$P^{(1)}(1) = \frac{p^2 + pq}{p^2} = \frac{p(p+q)}{p^2} = \frac{1}{p}.$$ ■

Como ya tenemos la función generadora de momento para ayudar a hallar los momentos de una variable aleatoria, ¿de qué valor es $P(t)$? La respuesta es que puede ser difícil encontrar $m(t)$ pero es mucho más fácil hallar $P(t)$. Entonces, $P(t)$ da una herramienta adicional para obtener los momentos de una variable aleatoria. Puede o no ser útil en una situación determinada.

Hallar los momentos de una variable aleatoria no es el principal uso de la función generadora de probabilidad. Su principal aplicación está en deducir la función de probabilidad (y por tanto la distribución de probabilidad) para otras variables aleatorias de valor entero relacionadas. Para estas aplicaciones, vea la obra de Feller (1968) y Parzen (1992).

Ejercicios

***3.164** Denote con Y una variable aleatoria binomial con n intentos y probabilidad de éxito p. Encuentre la función generadora de probabilidad para Y y úsela para hallar $E(Y)$.

***3.165** Denote con Y una variable aleatoria de Poisson con media λ. Encuentre la función generadora de probabilidad para Y y úsela para hallar $E(Y)$ y $V(Y)$.

***3.166** Consulte el Ejercicio 3.165. Use la función generadora de probabilidad hallada en él para encontrar $E(Y^3)$.

3.11 Teorema de Tchebysheff

Hemos visto en la Sección 1.3 y el Ejemplo 3.22 que si la probabilidad o histograma poblacional tiene forma aproximada de campana y se conocen la media y la varianza, la regla empírica es de gran ayuda para calcular las probabilidades de ciertos intervalos. No obstante, en muchos casos, las formas de los histogramas de probabilidad difieren marcadamente de la de un montículo y la regla empírica puede no dar aproximaciones útiles a las probabilidades de interés. El siguiente resultado, conocido como teorema de Tchebysheff, se puede usar para determinar un límite inferior para la probabilidad de que la variable aleatoria Y de interés caiga en un intervalo $\mu \pm k\sigma$.

TEOREMA **3.14** **Teorema de Tchebysheff** Sea Y una variable aleatoria con media μ y varianza finita σ^2. Entonces, para cualquier constante $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{o} \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Deben señalarse dos aspectos importantes de este resultado. Primero, el resultado se aplica a *cualquier* distribución de probabilidad, ya sea que el histograma de probabilidad tenga forma de campana o no. En segundo término, los resultados del teorema son muy conservadores en el sentido de que la probabilidad real de que Y esté en el intervalo $\mu \pm k\sigma$ por lo general excede del límite inferior para la probabilidad, $1 - 1/k^2$, por una cantidad considerable. No obstante, como dijimos en el Ejercicio 3.169, para cualquier $k > 1$, es posible construir una distribución de probabilidad tal que, para esa k, el límite provisto por el teorema de Tchebysheff se alcance en realidad. (El estudiante debe verificar que los resultados de la regla empírica no contradigan a los dados por el Teorema 3.14.) La prueba de este teorema se deja hasta la Sección 4.10. La utilidad de este teorema queda ilustrada en el siguiente ejemplo.

EJEMPLO 3.28 El número de clientes por día en un mostrador de ventas, Y, ha sido observado durante un largo periodo y se encontró que tiene una media de 20 y desviación estándar de 2. La distribución de probabilidad de Y no se conoce. ¿Qué se puede decir acerca de la probabilidad de que, mañana, Y sea mayor que 16 pero menor que 24?

Solución Deseamos hallar $P(16 < Y < 24)$. Del Teorema 3.14 sabemos que, para cualquier $k \geq 0$, $P(|Y - \mu| < k\sigma) \geq 1 - 1/k^2$ o,

$$P[(\mu - k\sigma) < Y < (\mu + k\sigma)] \geq 1 - \frac{1}{k^2}.$$

Como $\mu = 20$ y $\sigma = 2$, se deduce que $\mu - k\sigma = 16$ y $\mu + k\sigma = 24$ si $k = 2$. Entonces,

$$P(16 < Y < 24) = P(\mu - 2\sigma < Y < \mu + 2\sigma) \geq 1 - \frac{1}{(2)^2} = \frac{3}{4}.$$

En otras palabras, el total de clientes de mañana será entre 16 y 24 con una probabilidad más bien alta (al menos $3/4$).

Nótese que si σ fuera 1, k sería 4, y

$$P(16 < Y < 24) = P(\mu - 4\sigma < Y < \mu + 4\sigma) \geq 1 - \frac{1}{(4)^2} = \frac{15}{16}.$$

En consecuencia, el valor de σ tiene un efecto considerable sobre las probabilidades asociadas con intervalos. ■

Ejercicios

3.167 Sea Y una variable aleatoria con media 11 y varianza 9. Usando el Teorema de Tchebysheff, encuentre

a un límite inferior para $P(6 < Y < 16)$,

b el valor de C tal que $P(|Y - 11| \geq C) \leq .09$.

3.168 Qué preferiría, ¿hacer un examen de opción múltiple o uno ordinario? Si no sabe nada del material del examen, obtendrá una calificación de cero en uno ordinario. No obstante, si se le dan 5 opciones por cada pregunta de selección múltiple, tiene al menos una probabilidad en cinco de adivinar cada respuesta correcta. Suponga que un examen de opción múltiple contiene 100 preguntas, cada una con 5 posibles respuestas y supone la respuesta de cada una de las preguntas.

a ¿Cuál es el valor esperado del número Y de preguntas que serán contestadas correctamente?

b Encuentre la desviación estándar de Y.

c Calcule los intervalos $\mu \pm 2\sigma$ y $\mu \pm 3\sigma$.

d Si los resultados del examen son curvados de modo que 50 respuestas correctas ameritan una calificación de aprobación, ¿es probable que reciba una calificación aprobaria? Explique.

3.169 Este ejercicio demuestra que, en general, los resultados dados por el teorema de Tchebysheff no se pueden mejorar. Sea Y una variable aleatoria tal que

$$p(-1) = \frac{1}{18}, \qquad p(0) = \frac{16}{18}, \qquad p(1) = \frac{1}{18}.$$

a Demuestre que $E(Y) = 0$ y $V(Y) = 1/9$.

b Use la distribución de probabilidad de Y para calcular $P(|Y - \mu| \geq 3\sigma)$. Compare esta probabilidad exacta con el límite superior dado por el teorema de Tchebysheff para ver que el límite dado por dicho teorema se alcanza en realidad cuando $k = 3$.

***c** En el inciso b garantizamos que $E(Y) = 0$ al poner toda la masa de probabilidad en los valores -1, 0 y 1, con $p(-1) = p(1)$. La varianza fue controlada por las probabilidades asignadas a $p(-1)$ y $p(1)$. Usando esta misma idea básica, construya una distribución de probabilidad para una variable aleatoria X que dará $P(|X - \mu_X| \geq 2\sigma_X) = 1/4$.

***d** Si se especifica cualquier $k > 1$, ¿cómo puede construirse una variable aleatoria W de modo que $P(|W - \mu_W| \geq k\sigma_W) = 1/k^2$?

3.170 La Casa de Moneda de Estados Unidos produce monedas de diez centavos, las cuales tienen un diámetro promedio de .5 pulgadas y desviación estándar .01. Mediante el teorema de Tchebysheff encuentre un límite inferior para el número de monedas de un lote de 400 que se espera tengan un diámetro entre .48 y .52.

3.171 Para cierto tipo de suelo, el número de lombrices por pie cúbico tiene una media de 100. Suponiendo una distribución de Poisson de las lombrices, proponga un intervalo que incluirá al menos 5/9 de los valores muestrales de las cantidades de lombrices obtenida de un número grande de muestras de 1 pie cúbico.

3.172 Consulte el Ejercicio 3.115. Usando el histograma de probabilidad, encuentre la fracción de valores de la población que caiga a no más de 2 desviaciones estándar de la media. Compare su resultado con el del teorema de Tchebysheff.

3.173 Una moneda balanceada se lanza al aire tres veces. Sea Y igual al número de caras observado.

a Use la fórmula para la distribución de probabilidad binomial para calcular las probabilidades asociadas con $Y = 0, 1, 2$ y 3.

b Construya una distribución de probabilidad semejante a la de la Tabla 3.1.

c Encuentre el valor esperado y la desviación estándar de Y usando las fórmulas $E(Y) = np$ y $V(Y) = npq$.

d Usando la distribución de probabilidad del inciso b, encuentre la fracción de las medidas de población que se encuentren a no más de 1 desviación estándar de la media. Repita para 2 desviaciones estándar. ¿Cómo se comparan sus resultados con los del teorema de Tchebysheff y la regla empírica?

3.174 Suponga que una moneda estaba definitivamente no balanceada y que la probabilidad de una cara era igual a $p = .1$. Siga las instrucciones a, b, c y d que se expresan en el Ejercicio 3.173. Observe que la distribución de probabilidad pierde su simetría y se hace sesgada cuando p no es igual a 1/2.

3.175 En mayo de 2005, Tony Blair fue elegido para un histórico tercer mandato como primer ministro británico. Una encuesta de Gallup del Reino Unido (http://gallup.com/poll/content/default.aspx?ci=1710, junio 28 de 2005) realizada después de la elección de Blair indicó que a sólo 32% de los adultos británicos les gustaría ver a su hijo o hija ser primer ministro. Si la misma proporción de estadounidenses prefiriera que su hijo o hija fuera presidente y se hubieran entrevistado 120 adultos norteamericanos,

a ¿cuál es el número esperado de estadounidenses que preferirían que su hijo fuera presidente?

b ¿cuál es la desviación estándar del número Y que preferiría que su hijo fuera presidente?

c ¿es probable que exceda de 40 el número de estadounidenses que preferirían ver que su hijo fuera presidente?

3.176 Una encuesta nacional de 549 adolescentes (de 13 a 17 años de edad) realizada por Gallup (http://gallup.com/content/default.aspex?ci=17110), abril, 2005) indicó que 85% "pensaban que la ropa con símbolos de pandillas" debía prohibirse en las escuelas. Si en realidad los adolescentes estuvieran divididos por igual en sus opiniones respecto a la prohibición de ropa que lleve símbolos de pandillas, comente sobre la probabilidad de observar el resultado de usar esta encuesta (es decir, observar 85% o más en una muestra de 549 que estén a favor de prohibir ropa que lleve símbolos de pandillas). ¿Qué suposición debe hacerse acerca del procedimiento de muestreo para calcular esta probabilidad? [*Sugerencia:* recuerde el teorema de Tchebysheff y la regla empírica.]

3.177 Para cierta sección de un bosque de pinos, el número de árboles enfermos por acre, Y, tiene una distribución de Poisson con media $\lambda = 10$. Los árboles enfermos son rociados con un insecticida a un costo de \$3 por árbol, más un costo fijo general por renta de equipo de \$50. Si se denota con C el costo total de rociado para un acre seleccionado al azar, encuentre el valor esperado y la desviación estándar para C. ¿Dentro de qué intervalo espera el lector que se encuentre C con probabilidad de al menos .75?

3.178 Se sabe que 10% de una marca de tubos electrónicos para televisión se quema antes de que expire su garantía. Si se venden 1000 tubos, encuentre el valor esperado y la varianza de Y, el número de tubos originales que deben ser cambiados. ¿Dentro de qué límites se esperaría que caiga Y?

3.179 Consulte el Ejercicio 3.91. En éste, determinamos que la media y la varianza de los costos necesarios para hallar tres empleados con indicaciones positivas de envenenamiento por asbesto fueron 150 y 4500, respectivamente. ¿Piensa que es muy poco probable que el costo de completar las pruebas exceda de \$350?

3.12 Resumen

Este capítulo exploró las variables aleatorias discretas, sus distribuciones de probabilidad y sus valores esperados. El cálculo de la distribución de probabilidad para una variable aleatoria discreta requiere el uso de los métodos probabilísticos del Capítulo 2 para evaluar las probabilidades de eventos numéricos. Las funciones de probabilidad, $p(y) = P(Y = y)$, se obtuvieron para variables aleatorias binomiales, geométricas, binomiales negativas, hipergeométricas y de Poisson. Estas funciones de probabilidad a veces reciben el nombre de *funciones de masa de probabilidad*, porque dan la probabilidad (masa) asignada a cada uno de los posibles valores finitos o contablemente infinitos para estas variables aleatorias discretas.

Los valores esperados de variables aleatorias y funciones de variables aleatorias dieron un método para hallar la media y la varianza de Y y, en consecuencia, medidas de centralidad y variación para $p(y)$. Mucho del material restante en el capítulo se dedicó a las técnicas para adquirir expectativas, que a veces requerían sumar series aparentemente difíciles. Las técnicas para obtener expresiones de forma cerrada para algunos de los valores esperados resultantes incluyeron (1) usar el hecho de que $\sum_y p(y) = 1$ para cualquier variable aleatoria discreta y (2) $E(Y^2) = E[Y(Y-1)] + E(Y)$. Las medias y las varianzas de varias de las distribuciones discretas más comunes se resumen en la Tabla 3.4. Estos resultados y otros se encuentran también en la Tabla A2.1 del Apéndice 2 y al final de este libro.

La Tabla 3.5 proporciona los procedimientos *R* (y *S*-Plus) que dan $p(y_0) = P(Y = y_0)$ y $P(Y \leq y_0)$ para variables aleatorias con distribuciones binomiales, geométricas, binomiales negativas, hipergeométricas y de Poisson.

Posteriormente estudiamos la función generadora de momento asociada con una variable. Aun cuando a veces es útil para hallar μ y σ, la función generadora de momento es de valor primordial para que el estadístico teórico obtenga la distribución de probabilidad de una variable aleatoria. Las funciones generadoras de momento para casi todas las variables aleatorias comunes se encuentran en el Apéndice 2 y al final de este libro.

Tabla 3.4 Medias y varianzas para algunas variables aleatorias discretas comunes

Distribución	$E(Y)$	$V(Y)$
Binomial	np	$np(1-p) = npq$
Geométrica	$\frac{1}{p}$	$\frac{1-p}{p^2} = \frac{q}{p^2}$
Hipergeométrica	$n\left(\frac{r}{N}\right)$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$
de Poisson	λ	λ
Binomial negativa	$\frac{r}{p}$	$\frac{r(1-p)}{p^2} = \frac{rq}{p^2}$

Tabla 3.5 Procedimientos *R* (y *S*-Plus) que dan probabilidades para algunas distribuciones discretas comunes

Distribución	$P(Y = y_0) = p(y_0)$	$P(Y \leq y_0)$
Binomial	`dbinom(`y_0`,n,p)`	`pbinom(`y_0`,n,p)`
Geométrica	`dgeom(`y_0`-1,p)`	`pgeom(`y_0`-1,p)`
Hipergeométrica	`dhyper(`y_0`,r,N-r,n)`	`phyper(`y_0`,r,N-r,n)`
de Poisson	`dpois(`y_0`,`λ`)`	`ppois(`y_0`,`λ`)`
Binomial negativa	`dnbinom(`y_0`-r,r,p)`	`pnbinom(`y_0`-r,r,p)`

La función generadora de probabilidad es un procedimiento útil para obtener momentos y distribuciones de probabilidad de variables aleatorias de valores enteros.

Por último, dimos al teorema de Tchebysheff un resultado muy útil que permite calcular ciertas probabilidades cuando se conocen sólo la media y varianza.

Para concluir este resumen, recordemos el objetivo primordial de estadística: hacer una inferencia acerca de una población con base en información contenida en una muestra. Sacar la muestra de la población es el experimento. La muestra es en ocasiones un conjunto de medidas de una o más variables aleatorias y es el evento observado resultante de una sola repetición del experimento. Finalmente, hacer la inferencia acerca de la población requiere el conocimiento de la probabilidad de que ocurra la muestra observada, que a su vez requiere saber las distribuciones de probabilidad de las variables aleatorias que generaron la muestra.

Bibliografía y lecturas adicionales

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, 3d ed., vol. 1. New York: Wiley.

Goranson, U. G., and J. Hall. 1980. "Airworthiness of Long-Life Jet Transport Structures," *Aeronautical Journal* 84(838): 279–80.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Johnson, N. L., S. Kotz, and A. W. Kemp. 1993. *Univariate Discrete Distributions*, 2d ed. New York: Wiley.

Mosteller, F., R. E. K. Rourke, and G. B. Thomas. 1970. *Probability with Statistical Applications*, 2d ed. Reading, Mass. Addison-Wesley.

Parzen, E. 1964. *Stochastic Processes*. San Francisco: Holden-Day.

———. 1992.*Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Zwilliger, D. 2002. *CRC Standard Mathematical Tables*, 31st ed. Boca Raton, Fla.: CRC Press.

Ejercicios complementarios

3.180 Cuatro números posiblemente ganadores de una lotería —AB-4536, NH-7812, SQ-7855 y ZY-3221— llegan por correo. Usted gana un premio si uno de sus cuatro números es igual a uno de los números ganadores contenidos en una lista que tienen quienes dirigen la lotería. Se concede un primer premio de \$100,000, dos segundos premios de \$50,000 cada uno y diez terceros premios de \$1000 cada uno. Para ser elegible para ganar, es necesario enviar por correo el cupón a la compañía a un costo de 33¢ por envío. No se requiere hacer compras. De la estructura de los números que se reciban, es obvio que los números enviados constan de dos letras seguidas de cuatro dígitos. Suponiendo que los números recibidos se generaron al azar, ¿cuáles son las expectativas de que gane la lotería? ¿Merecen la pena los 33¢ para entrar a esta lotería?

3.181 El muestreo de piezas defectuosas de grandes lotes de productos manufacturados da un número de piezas defectuosas, Y, que sigue una distribución de probabilidad binomial. Un plan de muestreo consiste en especificar el número de piezas n por incluirse en una muestra y un número de aceptación a. El lote es aceptado si $Y \leq a$ y rechazado si $Y > a$. Denote con p la proporción de piezas defectuosas del lote. Para $n = 5$ y $a = 0$, calcule la probabilidad de aceptación del lote si (a) $p = 0$, (b) $p = .1$, (c) $p = .3$, (d) $p = .5$, (e) $p = 1.0$. Una gráfica que muestra la probabilidad de aceptación de lote como función de la fracción defectuosa del lote se llama *curva característica de operación* para el plan muestral. Construya la curva característica de operación para el plan $n = 5$, $a = 0$. Observe que un plan de muestreo es un ejemplo de inferencia estadística. Aceptar o rechazar un lote con base en información contenida en la muestra es equivalente a concluir que el lote es bueno o malo. "Bueno" implica que una parte pequeña es defectuosa y que por tanto el lote es apropiado para despacharse.

3.182 Consulte el Ejercicio 3.181. Use la Tabla 1, Apéndice 3 para construir las curvas características de operación para los siguientes planes de muestreo:

a $n = 10, a = 0$.

b $n = 10, a = 1$.

c $n = 10, a = 2$.

Para cada plan de muestreo, calcule P(aceptación de lote) para $p = 0$, .05 .1, .3, .5, y 1.0. Nuestra intuición sugiere que con el plan de muestreo (a) sería mucho menos probable aceptar lotes malos que los planes (b) y (c). Una comparación visual de las curvas características de operación confirmará esta conjetura intuitiva.

3.183 Un ingeniero de control de calidad desea estudiar planes de muestreo alternativos: $n = 5$, $a = 1$ y $n = 25$, $a = 5$. En una hoja de papel para graficar o milimétrico, construya las curvas características de operación para ambos planes, haciendo uso de probabilidades de aceptación en $p = .05$, $p = .10$, $p = .20$, $p = .30$ y $p = .40$ en cada caso.

a Si usted fuera un vendedor que produce lotes con una fracción defectuosa que va de $p = 0$ a $p = .10$, ¿cuál de los dos planes de muestreo preferiría?

b Si usted fuera un comprador que desea protegerse contra la aceptación de lotes con una fracción defectuosa que exceda de $p = .30$, ¿cuál de los dos planes de muestreo preferiría?

3.184 El comisionado de una ciudad dice que 80% de las personas que viven en la ciudad están a favor de la recolección de basura por contrato a una empresa privada y no de la recolección por empleados del municipio. Para probar lo dicho por el comisionado, se seleccionaron al azar 25 residentes de la ciudad, dando 22 que prefieren contratar una compañía privada.

a Si lo dicho por el comisionado es correcto, ¿cuál es la probabilidad de que la muestra contenga al menos 22 que prefieran contratar una compañía privada?

b Si lo dicho por el comisionado es correcto, ¿cuál es la probabilidad de que exactamente 22 prefieran contratar una compañía privada?

c Con base en observar 22 en una muestra de 25 que prefieren contratar una compañía privada, ¿qué concluye el lector acerca del dicho del comisionado de que 80% de los residentes de la ciudad prefieren contratar una compañía privada?

3.185 A veinte estudiantes se les pidió seleccionar un entero entre 1 y 10. Ocho de ellos escogieron ya sea 4, 5 o 6.

a Si los estudiantes hacen su selección de manera independiente y es tan probable que cada uno escoja uno u otro, ¿cuál es la probabilidad de que 8 o más seleccionen ya sea 4, 5 o 6?

b Habiendo observado ocho estudiantes que seleccionaron ya sea 4, 5 o 6, ¿qué conclusión se obtendría con base en la respuesta al inciso a?

3.186 Consulte los Ejercicios 3.67 y 3.68. Denote con Y el número de intento en el que se encuentra el primer solicitante con estudios de computación. Si cada entrevista cuesta \$30, encuentre el valor esperado y la varianza del costo total en que se incurre al entrevistar candidatos hasta hallar un solicitante con estudios avanzados de computación. ¿Dentro de qué límites se espera caigan los costos de la entrevista?

3.187 Considere el juego siguiente: un jugador lanza un dado no cargado repetidas veces hasta que obtiene 2, 3, 4, 5 o 6. En otras palabras, el jugador continúa tirando el dado mientras obtenga números 1. Cuando tira un número que no es 1, deja de tirar.

a ¿Cuál es la probabilidad de que el jugador tire el dado exactamente tres veces?

b ¿Cuál es el número esperado de tiros necesario para obtener el primer número que no sea 1?

c Si tira un número que no es 1 en el primer tiro, el jugador recibe un pago de \$1. De otro modo, el pago se duplica por cada 1 que el jugador tire antes de tirar otro que no sea 1. Entonces, el jugador recibe \$2 si tira un 1 seguido de otro que no sea 1; \$4 si tira dos números 1 seguidos de otro que no sea 1; \$8 si tira tres números 1 seguidos de otro que no sea 1, etcétera. En general, si hacemos que Y sea el número de tiros necesarios para obtener el primero que no sea 1, entonces el jugador tira $(Y - 1)$ números 1 antes de tirar el primero que no sea 1 y recibe un pago de 2^{Y-1} dólares. ¿Cuál es la cantidad que se espera pagar al jugador?

3.188 Si Y es una variable aleatoria binomial basada en n intentos y probabilidad de éxito p, demuestre que

$$P(Y > 1 \mid Y \geq 1) = \frac{1 - (1 - p)^n - np(1 - p)^{n-1}}{1 - (1 - p)^n}.$$

3.189 Un motor de arranque que se emplea en un vehículo espacial tiene un alto porcentaje de confiabilidad y tiene fama de arrancar en cualquier momento con probabilidad .99999. ¿Cuál es la probabilidad de que haya al menos una falla en los siguientes 10,000 arranques?

3.190 Consulte el Ejercicio 3.115. Encuentre μ, el valor esperado de Y, para la población teórica mediante el uso de la distribución de probabilidad obtenida en el Ejercicio 3.115. Encuentre la media muestral $\overline{y}$ para las $n = 100$ mediciones generadas en el Ejercicio 3.116. ¿$\overline{y}$ da una buena estimación de μ?

3.191 Encuentre la varianza de la población σ^2 para el Ejercicio 3.115 y la varianza muestral s^2 para el Ejercicio 3.116. Compare.

3.192 Tire un dado balanceado y sea Y el número de puntos observados en la cara superior. Encuentre la media y la varianza de Y. Construya un histograma de probabilidad y localice el intervalo $\mu \pm 2\sigma$. Verifique que se cumpla el teorema de Tchebysheff.

3.193 Dos líneas de ensamble I y II tienen el mismo porcentaje de piezas defectuosas en su producción de reguladores de voltaje. Cinco reguladores se toman como muestra y se prueban en cada línea. Entre el total de diez reguladores probados, cuatro están defectuosos. Encuentre la probabilidad de que exactamente dos de los reguladores defectuosos provengan de la línea I.

3.194 La preocupación de una jugadora es quedar en bancarrota antes de lograr su primer triunfo. Suponga que ella juega una partida en la que la probabilidad de ganar es .1 (y ella lo desconoce). Le cuesta \$10 jugar y ella recibe \$80 por ganar. Si comienza con \$30, ¿cuál es la probabilidad de que gane exactamente una vez antes de perder su capital inicial?

3.195 El número de imperfecciones en el tejido de cierto textil tiene una distribución de Poisson con una media de 4 por yarda cuadrada. Encuentre la probabilidad de que

a una muestra de 1 yarda cuadrada contenga al menos una imperfección,

b una muestra de 3 yardas cuadradas contenga al menos una imperfección.

3.196 Consulte el Ejercicio 3.195. El costo de reparar las imperfecciones en el tejido es \$10 por imperfección. Encuentre la media y la desviación estándar del costo de reparación para un rollo de 8 yardas cuadradas del textil.

3.197 El número de colonias de cierto tipo de bacterias en muestras de agua contaminada tiene una distribución de Poisson con una media de 2 por centímetro cúbico (cm^3).

a Si cuatro muestras de 1 cm^3 se seleccionan de esta agua de manera independiente, encuentre la probabilidad de que al menos una contenga una o más colonias de bacterias.

b ¿Cuántas muestras de 1 cm^3 deben seleccionarse para tener una probabilidad de aproximadamente .95 de ver al menos una colonia de bacterias?

3.198 Un modelo para competencia de plantas supone que hay una zona de agotamiento de recursos alrededor de cada planta. Dependiente del tamaño de las zonas y la densidad de las plantas, las zonas de agotamiento de recursos pueden traslaparse con las de otras plantas en las cercanías. Cuando las semillas se dispersan al azar sobre un lugar amplio, el número de vecinos que cada planta tiene dentro de un área de tamaño A suele seguir una distribución de Poisson con media igual a $A \times d$, donde d es la densidad de plantas por unidad de área. Suponga que la densidad de plantas es cuatro por metro cuadrado. ¿Cuál es la probabilidad de que una planta especificada

a no tenga vecinos en no más de un metro?,

b tenga al menos tres vecinos en no más de 2 metros?

3.199 La diabetes dependiente de insulina (IDD, por sus siglas en inglés) es una enfermedad crónica común en niños. La enfermedad se presenta con más frecuencia en niños descendientes de europeos del norte, pero la incidencia va de una baja de 1–2 casos por 100,000 por año a una alta de más de 40 casos por 100,000 en algunas partes de Finlandia.[4] Supongamos que una región en Europa tiene una incidencia de 30 casos por 100,000 por año y que al azar seleccionamos 1000 niños de esta región.

a ¿La distribución del número de casos de IDD entre los de la muestra puede ser aproximada por una distribución de Poisson? Si es así, ¿cuál es la media de la aproximación de distribución de Poisson?

b ¿Cuál es la probabilidad de que observemos al menos dos casos de IDD entre los 1000 niños de la muestra?

3.200 Usando el dato de que

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots,$$

expanda la función generadora de momento para la distribución binomial

$$m(t) = (q + pe^t)^n$$

en una serie de potencias en t. (Tome en cuenta sólo los términos de orden inferior en t.) Identifique μ_i' como el coeficiente de $t^i/i!$ que aparece en la serie. Específicamente, encuentre μ_1' y μ_2' y compárelas con los resultados del Ejercicio 3.146.

3.201 Consulte los Ejercicios 3.103 y 3.106. ¿En qué intervalo esperaría encontrar los costos de reparación en estas cinco máquinas? (Use el teorema de Tchebysheff.)

***3.202** El número de autos que pasan por una zona de estacionamiento en un intervalo de un minuto tiene una distribución de Poisson con media λ. La probabilidad de que cualquier conductor individual desee estacionar su auto es p. Suponga que las personas decidan estacionarse de modo independiente entre sí.

a Si hay un lugar de estacionamiento y tomará 1 minuto llegar ahí, ¿cuál es la probabilidad de que el lugar esté disponible cuando la persona llegue al lote? (Suponga que nadie sale del lote durante el intervalo de un minuto.)

b Denote con W el número de conductores que deseen estacionarse durante un intervalo de un minuto. Deduzca la distribución de probabilidad de W.

3.203 Un tipo de célula de bacteria se divide a un ritmo constante de λ en el tiempo. (Es decir, la probabilidad de que una célula se divida en un intervalo de t más corto es aproximadamente λt.) Dado que una población inicia en el tiempo cero con k células de esta bacteria y que las divisiones celulares son independientes entre sí, el tamaño de la población en el tiempo t, $Y(t)$, tiene una distribución de probabilidad

$$P[Y(t) = n] = \binom{n-1}{k-1} e^{-\lambda k t}\left(1 - e^{-\lambda t}\right)^{n-k}, \qquad n = k,\ k+1, \ldots$$

a Encuentre el valor esperado y la varianza de $Y(t)$ en términos de λ y t.

b Si, para un tipo de célula de bacteria, $\lambda = .1$ por segundo y la población empieza con dos células en el tiempo cero, encuentre el valor esperado y la varianza de la población después de cinco segundos.

3.204 La probabilidad de que cualquier conductor de automóvil gire a la izquierda en un crucero es .2. El carril de vuelta a la izquierda en este crucero tiene espacio para tres vehículos. Si el carril de giro a la izquierda está vacío cuando el semáforo se pone en rojo y cinco vehículos llegan a este crucero cuando la luz está en rojo, encuentre la probabilidad de que el carril de vuelta a la izquierda contenga los vehículos de todos los conductores que deseen dar vuelta a la izquierda.

3.205 Un experimento consiste en tirar un dado no cargado hasta que un 6 salga cuatro veces. ¿Cuál es la probabilidad de que el proceso termine después de exactamente diez tiros con un 6 saliendo en los tiros noveno y décimo?

4. M. A. Atkinson, "Diet, Genetics, and Diabetes", en *Food Technology* 51(3), (1997): 77.

3.206 Los registros de accidentes recolectados por una compañía de seguros de automóviles dan la siguiente información. La probabilidad de que un conductor asegurado tenga un accidente automovilístico es .15. Si ha ocurrido un accidente, las averías al vehículo ascienden al 20% de su valor comercial con una probabilidad de .80, a 60% de su valor comercial con una probabilidad de .12 y a pérdida total con una probabilidad de .08. ¿Qué prima debe cobrar la compañía sobre un auto de \$12,000 para que la ganancia esperada por la compañía sea cero?

3.207 El número de personas que entran en la unidad de cuidados intensivos de un hospital en cualquier día posee una distribución de Poisson con media igual a cinco personas por día.

a ¿Cuál es la probabilidad de que el número de personas que entren en la unidad de cuidados intensivos en un día particular sea igual a 2? ¿y que sea menor o igual que 2?

b ¿Es probable que Y pase de 10? Explique.

3.208 Una encuesta reciente sugiere que los estadounidenses anticipan una reducción en sus estándares de vida y que un nivel cada vez más creciente de consumo ya no es tan importante como lo fue en el pasado. Suponga que una encuesta de 2000 personas indica que 1373 están a favor de que, por medios legislativos, se obligue a una reducción en el tamaño de los automóviles hechos en Estados Unidos. ¿Esperaría observar hasta 1373 personas a favor de esta proposición si, de hecho, el público general se ha dividido en 50 a 50 con respecto a este problema? ¿Por qué?

3.209 Un proveedor de maquinaria pesada para construcción ha encontrado que normalmente se obtienen nuevos clientes por medio de solicitudes de éstos en llamadas de ventas y que la probabilidad de venta de una pieza particular de equipo es .3. Si el proveedor tiene tres máquinas para venta, ¿cuál es la probabilidad de que se requieran menos de cinco llamadas de clientes para agotar el inventario?

3.210 Calcule $P(|Y - \lambda| \leq 2\sigma)$ para la distribución de probabilidad de Poisson del Ejemplo 3.22. ¿Esto está de acuerdo con la regla empírica?

***3.211** Una comerciante tiene en existencia cierto artículo perecedero. Ella sabe que en cualquier día determinado tendrá una demanda de dos, tres o cuatro de estos artículos con probabilidades .1, .4 y .5, respectivamente. Compra los artículos en \$1.00 cada uno y los vende en \$1.20 cada uno. Si quedan algunos al final del día, representan pérdida total. ¿Cuántos artículos debe tener en existencia para maximizar su utilidad diaria esperada?

***3.212** Demuestre que la función de probabilidad hipergeométrica se aproxima a la binomial en el límite cuando $N \to \infty$ y $p = r/N$ permanece constante. Esto es, demuestre que

$$\lim_{N\to\infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y q^{n-y},$$

para $p = r/N$ constante.

3.213 Un lote de $N = 100$ productos industriales contiene 40 defectuosos. Sea Y el número de defectuosos en una muestra aleatoria de tamaño 20. Encuentre $p(10)$ con el uso de (a) la distribución de probabilidad hipergeométrica y (b) la distribución de probabilidad binomial. ¿Es N suficientemente grande para que el valor de $p(10)$ obtenido con de la distribución binomial sea una buena aproximación de la obtenida usando la distribución hipergeométrica?

***3.214** Por simplicidad, supongamos que hay dos clases de conductores. Los conductores seguros, que son 70% de la población, tienen probabilidad .1 de causar un accidente en un año. El resto de la población son provocadores de accidentes, que tienen una probabilidad .5 de causar un accidente en un año. La prima de seguro es \$400 multiplicado por la probabilidad de alguien de causar un accidente en el año siguiente. Un nuevo suscriptor tiene un accidente durante el primer año. ¿Cuál debe ser la prima de seguro para el año siguiente?

***3.215** Se sabe que 5% de los miembros de una población tienen una enfermedad A, que puede ser descubierta por un examen de sangre. Suponga que N (un número grande) personas se someten al examen. Esto puede hacerse en dos formas: (1) Cada persona es examinada por separado o (2) las muestras sanguíneas de k personas se agrupan y analizan. (Suponga que $N = nk$, con n un entero.) Si el examen es negativo, todos ellos son sanos (es decir, sólo se requiere este examen). Si es positivo, cada una de las k personas debe ser examinada por separado (es decir, se requiere un total de $k + 1$ exámenes).

a Para una k fija, ¿cuál es el número esperado de exámenes necesario en la opción 2?

b Encuentre la k que minimizará el número esperado de exámenes en la opción 2.

c Si k se selecciona como en el inciso b, ¿en promedio cuántos exámenes ahorra la opción 2 en comparación con la opción 1?

***3.216** Considere que Y tiene una distribución hipergeométrica

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, \qquad y = 0,\ 1,\ 2, \ldots,\ n.$$

a Demuestre que

$$P(Y = n) = p(n) = \left(\frac{r}{N}\right)\left(\frac{r-1}{N-1}\right)\left(\frac{r-2}{N-2}\right)\cdots\left(\frac{r-n+1}{N-n+1}\right).$$

b Escriba $p(y)$ como $p(y|r)$. Demuestre que si $r_1 < r_2$, entonces

$$\frac{p(y|r_1)}{p(y|r_2)} > \frac{p(y+1|r_1)}{p(y+1|r_2)}.$$

c Aplique la expansión binomial a cada factor de la siguiente ecuación:

$$(1+a)^{N_1}(1+a)^{N_2} = (1+a)^{N_1+N_2}.$$

Ahora compare los coeficientes de a^n en ambos lados para demostrar que

$$\binom{N_1}{0}\binom{N_2}{n} + \binom{N_1}{1}\binom{N_2}{n-1} + \cdots + \binom{N_1}{n}\binom{N_2}{0} = \binom{N_1+N_2}{n}.$$

d Usando el resultado del inciso c, concluya que

$$\sum_{y=0}^{n} p(y) = 1.$$

***3.217** Use el resultado obtenido en el Ejercicio 3.216 c y la Definición 3.4 para obtener directamente la media de una variable aleatoria hipergeométrica.

***3.218** Use los resultados de los Ejercicios 3.216 y 3.217 para demostrar que, para una variable aleatoria hipergeométrica,

$$E[Y(Y-1)] = \frac{r(r-1)n(n-1)}{N(N-1)}.$$

CAPÍTULO 4

Variables continuas y sus distribuciones de probabilidad

4.1 Introducción

Un momento de reflexión sobre variables aleatorias que se encuentran en el mundo real debe convencernos de que no todas las variables aleatorias de interés son discretas. El número de días que llueve en un periodo de n días es una variable aleatoria discreta porque el número de días debe tomar uno de los $n + 1$ valores 0, 1, 2, . . . , o n. Ahora considere la lluvia diaria en un punto geográfico específico. En teoría, con equipo de medición de precisión perfecta, la cantidad de lluvia podría tomar cualquier valor entre 0 y 5 pulgadas. En consecuencia, cada uno del número incontable e infinito de puntos del intervalo (0, 5) representa un valor posible distinto de la cantidad de lluvia en un día. Una variable aleatoria que puede tomar cualquier valor en un intervalo se denomina *continua* y el propósito de este capítulo es estudiar distribuciones de probabilidad para variables aleatorias continuas. La producción de un antibiótico en un proceso de fermentación es una variable aleatoria continua, al igual que la duración de vida útil, en años, de una máquina lavadora. Los segmentos de recta sobre los cuales estas dos variables se definen están contenidos en la mitad positiva de la recta real. Esto no significa que, si observamos suficientes máquinas lavadoras, podríamos a la larga observar un resultado correspondiente a cada valor del intervalo (3, 7); más bien, esto significa que ningún valor entre 3 y 7 se puede excluir como posible para el número de años que una máquina lavadora permanece en servicio.

La distribución de probabilidad para una variable aleatoria discreta siempre puede darse al asignar una probabilidad no negativa a cada uno de los posibles valores que la variable pueda tomar. En todo caso, por supuesto, la suma de todas las probabilidades que asignamos debe ser igual a 1. Por desgracia, la distribución de probabilidad para una variable aleatoria continua no puede ser especificada en la misma forma. Es matemáticamente imposible asignar probabilidades diferentes de cero a todos los puntos de un intervalo de recta al tiempo que se satisface el requisito de que las probabilidades de los distintos valores posibles asciendan a 1. En consecuencia, debemos desarrollar un método diferente para describir la distribución de probabilidad para una variable aleatoria continua.

4.2 Distribución de probabilidad para una variable aleatoria continua

Antes de que podamos expresar una definición formal para una variable aleatoria continua, debemos definir la función de distribución (o función de distribución *acumulativa*) asociada con una variable aleatoria.

DEFINICIÓN 4.1 Denote con Y cualquier variable aleatoria. La *función de distribución* de Y, denotada por $F(y)$, es tal que $F(y) = P(Y \leq y)$ para $-\infty < y < \infty$.

La naturaleza de la función de distribución asociada con una variable aleatoria determina si la variable es continua o discreta. En consecuencia, comenzaremos nuestra exposición al examinar la función de distribución para una variable aleatoria discreta y tomando nota de las características de esta función.

EJEMPLO 4.1 Suponga que Y tiene una distribución binomial con $n = 2$ y $p = 1/2$. Encuentre $F(y)$.

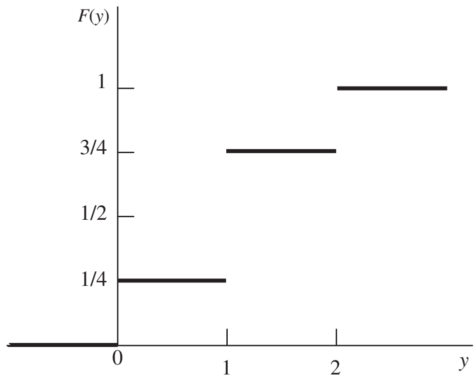
Solución La función de probabilidad para Y está dada por

$$p(y) = \binom{2}{y}\left(\frac{1}{2}\right)^{y}\left(\frac{1}{2}\right)^{2-y}, \qquad y = 0,\ 1,\ 2,$$

y se obtiene

$$p(0) = 1/4, \qquad p(1) = 1/2, \qquad p(2) = 1/4.$$

FIGURA **4.1**
Función de distribución binomial, $n = 2$, $p = 1/2$



¿Cuál es $F(-2) = P(Y \leq -2)$? Como los únicos valores de Y a los que se asignan probabilidades positivas son 0, 1 y 2 y ninguno de estos valores son menores o iguales a –2, $F(-2) = 0$. Si usamos una lógica similar, $F(y) = 0$ para toda $y < 0$. ¿Cuál es $F(1.5)$? Los únicos valores de Y que son menores o iguales a 1.5 y tienen probabilidades diferentes de cero son los valores 0 y 1. Por lo que,

$$\begin{aligned} F(1.5) = P(Y \leq 1.5) &= P(Y = 0) + P(Y = 1) \\ &= (1/4) + (1/2) = 3/4. \end{aligned}$$

En general,

$$F(y) = P(Y \leq y) = \begin{cases} 0, & \text{para } y < 0, \\ 1/4, & \text{para } 0 \leq y < 1, \\ 3/4, & \text{para } 1 \leq y < 2, \\ 1, & \text{para } y \geq 2. \end{cases}$$

Una gráfica de $F(y)$ se da en la Figura 4.1. ■

En el Ejemplo 4.1, los puntos entre 0 y 1 o entre 1 y 2 tenían todos probabilidad 0 y no contribuyeron en nada a la probabilidad acumulativa descrita por la función de distribución. En consecuencia, la función de distribución acumulativa siguió siendo plana entre los posibles valores de Y y aumentó en saltos o escalones en cada uno de los posibles valores de Y. Las funciones que se comportan de ese modo se denominan *funciones escalón. Las funciones de distribución para variables aleatorias discretas son siempre funciones escalón porque la función de distribución acumulativa aumenta sólo en el número finito o contable de puntos con probabilidades positivas.*

Como la función de distribución asociada con cualquier variable aleatoria es tal que $F(y) = P(Y \leq y)$, desde un punto de vista práctico es evidente que $F(-\infty) = \text{lím}_{y\to-\infty} P(Y \leq y)$ debe ser cero. Si consideramos dos valores cualesquiera $y_1 < y_2$, entonces $P(Y \leq y_1) \leq P(Y \leq y_2)$, es decir, $F(y_1) \leq F(y_2)$. Entonces, una función de distribución, $F(y)$, es siempre una función monotónica, no decreciente. Además, es evidente que $F(\infty) = \text{lím}_{y\to\infty} P(Y \leq y) = 1$. Estas tres características definen las propiedades de cualquier función de distribución y están resumidas en el siguiente teorema.

TEOREMA **4.1**

Propiedades de una función de distribución[1] Si $F(y)$ es una función de distribución, entonces

1. $F(-\infty) \equiv \lim_{y \to -\infty} F(y) = 0$.
2. $F(\infty) \equiv \lim_{y \to \infty} F(y) = 1$.
3. $F(y)$ es una función no decreciente de y. [Si y_1 y y_2 son *cualesquiera* valores de manera que $y_1 < y_2$, entonces $F(y_1) \leq F(y_2)$.]

Usted deberá comprobar que la función de distribución desarrollada en el Ejemplo 4.1 tenga cada una de estas propiedades.

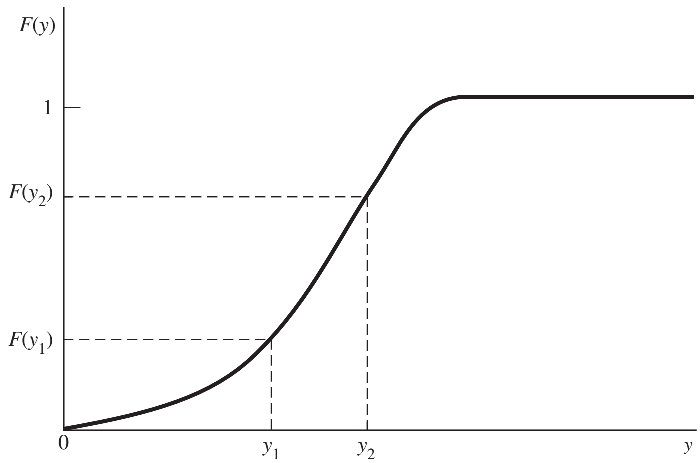
Examinemos ahora la función de distribución para una variable aleatoria continua. Suponga que, para todos los fines prácticos, la cantidad de lluvia diaria, Y, debe ser menor que 6 pulgadas. Para todo $0 \leq y_1 < y_2 \leq 6$, el intervalo (y_1, y_2) tiene una probabilidad positiva de incluir Y, sin importar cuánto se acerque y_1 a y_2. Se deduce que $F(y)$ en este caso debe ser una función lisa, creciente sobre algún intervalo de números reales, como se grafica en la Figura 4.2.

Por tanto, llegamos a la definición de una variable aleatoria continua.

DEFINICIÓN **4.2**

Una variable aleatoria Y con función de distribución $F(y)$ se dice que es *continua* si $F(y)$ es continua, para $-\infty < y < \infty$.[2]

FIGURA **4.2**
Función de distribución para una variable aleatoria continua



1. Para ser matemáticamente rigurosos, si $F(y)$ es una función de distribución válida, entonces $F(y)$ también debe ser continua.

2. Para ser matemáticamente precisos, también necesitamos que exista la primera derivada de $F(y)$ y que sea continua excepto para, a lo sumo, un número finito de puntos en cualquier intervalo finito. Las funciones de distribución para las variables aleatorias continuas estudiadas en este texto satisfacen este requisito.

Si y es una variable aleatoria continua, entonces, para cualquier número real y,

$$P(Y = y) = 0.$$

Si esto no fuera cierto y $P(Y = y_0) = p_0 > 0$, entonces $F(y)$ tendría una discontinuidad (salto) de tamaño p_0 en el punto y_0, violando la suposición de que Y era continua. Hablando en términos prácticos, el hecho de que las variables aleatorias continuas tengan probabilidad cero en puntos discretos no debe molestarnos. Considere el ejemplo de medir la lluvia diaria. ¿Cuál es la probabilidad de que veamos una medida de lluvia diaria de exactamente 2.193 pulgadas? Es bastante probable que nunca observemos ese valor exacto incluso si tomamos medidas de lluvia durante toda una vida, aunque podríamos ver muchos días con medidas entre 2 y 3 pulgadas.

La derivada de $F(y)$ es otra función de gran importancia en teoría de probabilidad y estadística.

DEFINICIÓN **4.3**

Sea $F(y)$ la función de distribución para una variable aleatoria continua Y. Entonces $f(y)$, dada por

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

siempre que exista la derivada, se denomina *función de densidad de probabilidad* para la variable aleatoria Y.

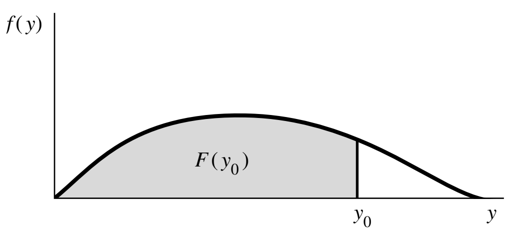
Se deduce de las Definiciones 4.2 y 4.3 que $F(y)$ se puede escribir como

$$F(y) = \int_{-\infty}^{y} f(t)\ dt,$$

donde $f(\cdot)$ es la función de densidad de probabilidad y t se usa como la variable de integración. La relación entre las funciones de distribución y de densidad se muestra gráficamente en la Figura 4.3.

La función de densidad de probabilidad es un *modelo teórico* para la distribución de frecuencia (histograma) de una población de medidas. Por ejemplo, observaciones de la vida útil de lavadoras de una marca particular generan mediciones que pueden estar caracterizadas por un histograma de frecuencia relativo, como se explica en el Capítulo 1. De manera conceptual, el experimento podría repetirse hasta el infinito, con lo cual se genera una distribución de frecuencia relativa (una curva suave) que caracterizaría la población de interés para el fabricante. Esta distribución teórica de frecuencia relativa corresponde a la función de densidad de probabilidad para la duración de vida de una sola máquina, Y.

FIGURA **4.3** La función de distribución



Debido a que la función de distribución $F(y)$ para cualquier variable aleatoria siempre tiene las propiedades dadas en el Teorema 4.1, las funciones de densidad deben tener algunas propiedades correspondientes. Como $F(y)$ es una función no decreciente, la derivada $f(y)$ nunca es negativa. Además, sabemos que $F(\infty) = 1$ y, por tanto, que $\int_{-\infty}^{\infty} f(t)\,dt = 1$. En resumen, las propiedades de una función de densidad de probabilidad se dan en el siguiente teorema.

TEOREMA **4.2**

Propiedades de una función de densidad Si $f(y)$ es una función de densidad para una variable aleatoria continua, entonces

1. $f(y) \geq 0$ para toda y, $-\infty < y < \infty$.
2. $\int_{-\infty}^{\infty} f(y)\,dy = 1$.

El siguiente ejemplo proporciona la función de distribución y la función de densidad para una variable aleatoria continua.

EJEMPLO **4.2** Suponga que

$$F(y) = \begin{cases} 0, & \text{para } y < 0, \\ y, & \text{para } 0 \leq y \leq 1, \\ 1, & \text{para } y > 1. \end{cases}$$

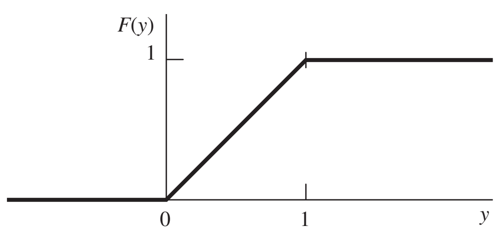
Encuentre la función de densidad de probabilidad para Y y grafíquela.

Solución Como la función de densidad $f(y)$ es la derivada de la función de distribución $F(y)$, cuando la derivada existe,

$$f(y) = \frac{dF(y)}{dy} = \begin{cases} \dfrac{d(0)}{dy} = 0, & \text{para } y < 0, \\ \dfrac{d(y)}{dy} = 1, & \text{para } 0 < y < 1, \\ \dfrac{d(1)}{dy} = 0, & \text{para } y > 1, \end{cases}$$

y $f(y)$ no está definida en $y = 0$ y $y = 1$. Una gráfica de $F(y)$ se muestra en la Figura 4.4.

FIGURA **4.4** Función de distribución $F(y)$ para el Ejemplo 4.2



■

La gráfica de $f(y)$ para el Ejemplo 4.2 se muestra en la Figura 4.5. Observe que las funciones de distribución y densidad dadas en el Ejemplo 4.2 tienen todas las propiedades requeridas

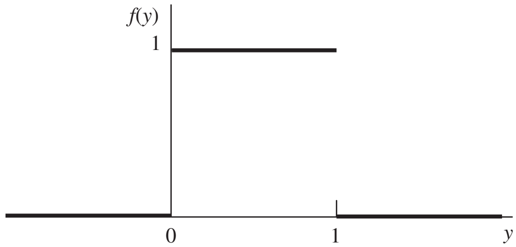


FIGURA **4.5** Función de densidad $f(y)$ para el Ejemplo 4.2

de las funciones de distribución y densidad, respectivamente. Además, $F(y)$ es una función continua de y, pero $f(y)$ es discontinua en los puntos $y = 0, 1$. En general, la función de distribución para una variable aleatoria continua debe ser continua, pero la función de densidad no necesita ser continua en todas partes.

EJEMPLO **4.3** Sea Y una variable aleatoria continua con función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $F(y)$. Grafique $f(y)$ y $F(y)$.

Solución La gráfica de $f(y)$ aparece en la Figura 4.6. Como

$$F(y) = \int_{-\infty}^{y} f(t)\,dt,$$

tenemos, para este ejemplo,

$$F(y) = \begin{cases} \int_{-\infty}^{y} 0\,dt = 0, & \text{para } y < 0, \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{y} 3t^2\,dt = 0 + t^3\big]_0^y = y^3, & \text{para } 0 \le y \le 1, \\ \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} 3t^2\,dt + \int_{1}^{y} 0\,dt = 0 + t^3\big]_0^1 + 0 = 1, & \text{para } 1 < y. \end{cases}$$

Observe que algunas de las integrales que hemos evaluado conducen a un valor de 0. Éstas se incluyen para hacer más completo este ejemplo inicial. En cálculos futuros no mostraremos de manera explícita ninguna integral que tenga valor 0. La gráfica de $F(y)$ se da en la Figura 4.7.

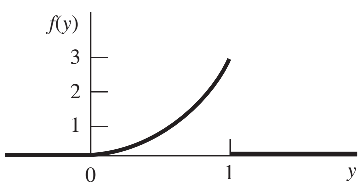


FIGURA **4.6** Función de densidad para el Ejemplo 4.3

■

$F(y_0)$ da la probabilidad de que $Y \le y_0$. Como se verá en los capítulos siguientes, en ocasiones estamos interesados en determinar el valor, y, de una variable aleatoria Y de manera que $P(Y \le y)$ iguale o exceda algún valor especificado.

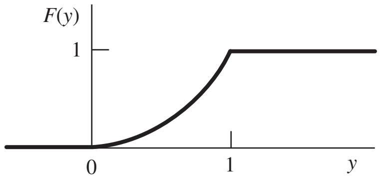


FIGURA 4.7 Función de distribución para el Ejemplo 4.3

DEFINICIÓN 4.4

Denotemos con Y cualquier variable aleatoria. Si $0 < p < 1$, el p–ésimo *cuantil* de Y, denotado por ϕ_p, es el mínimo valor tal que $P(Y \leq \phi_q) = F(\phi_p) \geq p$. Si Y es continua, ϕ_p es el mínimo valor tal que $F(\phi_p) = P(Y \leq \phi_q) = p$. Algunos prefieren llamar ϕ_p al $100p$–ésimo *percentil* de Y.

Un caso especial importante es $p = 1/2$ y $\phi_{.5}$ es la *mediana* de la variable aleatoria Y. En el Ejemplo 4.3 la mediana de la variable aleatoria es tal que $F(\phi_{.5}) = .5$ y fácilmente se ve, que $(\phi_{.5})^3 = .5$, o bien, de manera similar, que la mediana de Y es $\phi_{.5} = (.5)^{1/3} = .7937$.

El siguiente paso es hallar la probabilidad de que Y caiga en un intervalo específico; esto es, $P(a \leq Y \leq b)$. Del Capítulo 1 sabemos que esta probabilidad corresponde al área bajo la distribución de frecuencia en el intervalo $a \leq y \leq b$. Como $f(y)$ es la similar teórica de la distribución de frecuencia, podríamos esperar que $P(a \leq Y \leq b)$ fuera igual a un área correspondiente bajo la función de densidad $f(y)$. Esto de hecho es verdad porque, si $a < b$,

$$P(a < Y \leq b) = P(Y \leq b) - P(Y \leq a) = F(b) - F(a) = \int_a^b f(y)\, dy.$$

Como $P(Y = a) = 0$, tenemos el siguiente resultado.

TEOREMA 4.3

Si la variable aleatoria Y tiene función de densidad $f(y)$ y $a < b$, entonces la probabilidad de que Y caiga en el intervalo $[a, b]$ es

$$P(a \leq Y \leq b) = \int_a^b f(y)\, dy.$$

Esta probabilidad es el área sombreada de la Figura 4.8.

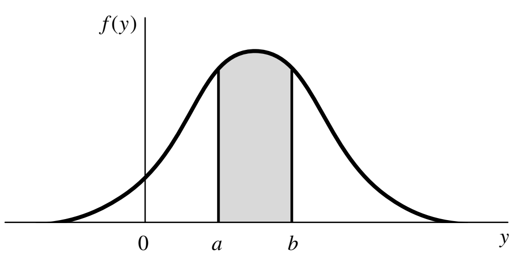


FIGURA 4.8 $P(a \leq Y \leq b)$

Si Y es una variable aleatoria continua y a y b son constantes tales que $a < b$, entonces $P(Y = a) = 0$ y $P(Y = b) = 0$ y el Teorema 4.3 implica que

$$\begin{aligned} P(a < Y < b) &= P(a \le Y < b) = P(a < Y \le b) \\ &= P(a \le Y \le b) = \int_a^b f(y)\, dy. \end{aligned}$$

El hecho de que esta secuencia de igualdades *no* es, en general, verdadera para variables aleatorias discretas se ilustra en el Ejercicio 4.7.

EJEMPLO **4.4** Dada $f(y) = cy^2$, $0 \le y \le 2$ y $f(y) = 0$ en cualquier otra parte, encuentre el valor de c para el cual $f(y)$ es una función de densidad válida.

Solución Requerimos un valor para c de manera que

$$\begin{aligned} F(\infty) = \int_{-\infty}^{\infty} f(y)\, dy &= 1 \\ &= \int_0^2 cy^2\, dy = \left.\frac{cy^3}{3}\right]_0^2 = \left(\frac{8}{3}\right)c. \end{aligned}$$

Entonces, $(8/3)c = 1$, y encontramos que $c = 3/8$. ■

EJEMPLO **4.5** Encuentre $P(1 \le Y \le 2)$ para el Ejemplo 4.4. También encuentre $P(1 < Y < 2)$.

Solución

$$P(1 \le Y \le 2) = \int_1^2 f(y)\, dy = \frac{3}{8}\int_1^2 y^2\, dy = \left(\frac{3}{8}\right)\left.\frac{y^3}{3}\right]_1^2 = \frac{7}{8}.$$

Como Y tiene una distribución continua, se deduce que $P(Y = 1) = P(Y = 2) = 0$ y, por tanto, que

$$P(1 < Y < 2) = P(1 \le Y \le 2) = \frac{3}{8}\int_1^2 y^2\, dy = \frac{7}{8}.$$ ■

Los enunciados de probabilidad que se refieren a una variable aleatoria continua Y son significativos sólo si, primero, existe la integral que define la probabilidad y, segundo, que las probabilidades resultantes concuerden con los axiomas del Capítulo 2. Estas dos condiciones siempre quedarán satisfechas si consideramos sólo probabilidades asociadas con un conjunto finito o contable de intervalos. Como casi siempre estamos interesados en probabilidades en las que las variables continuas caen en intervalos, esta consideración no nos ocasionará dificultad. Algunas funciones de densidad que dan buenos modelos para distribuciones de frecuencia poblacional, que se encuentran en aplicaciones prácticas, se presentan en secciones posteriores.

Ejercicios

4.1 Sea Y una variable aleatoria con $p(y)$ dada en la tabla siguiente.

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

a Obtenga la función de distribución, $F(y)$. Asegúrese de especificar el valor de $F(y)$ para toda y, $-\infty < y < \infty$.

b Trace la función de distribución dada en el inciso a.

4.2 Una caja contiene cinco llaves, sólo una de las cuales abrirá una cerradura. Las llaves se seleccionan al azar y se prueban una a la vez hasta que la cerradura se abre (las llaves que no funcionan se descartan antes de probar otra). Sea Y el número de intentos en los que la cerradura se abre.

a Encuentre la función de probabilidad para Y.

b Obtenga la correspondiente función de distribución.

c ¿Qué es $P(Y < 3)$? ¿$P(Y \leq 3)$? ¿$P(Y = 3)$?

d Si Y es una variable aleatoria continua, decimos que, para toda $-\infty < a < \infty$, $P(Y = a) = 0$. ¿Alguna de las respuestas en el inciso c contradice esta afirmación? ¿Por qué?

4.3 Una variable aleatoria de *Bernoulli* es aquella que toma sólo dos valores, 0 y 1 con $p(1) = p$ y $p(0) = 1 - p \equiv q$.

a Trace la correspondiente función de distribución.

b Demuestre que esta función de distribución tiene las propiedades dadas en el Teorema 4.1.

4.4 Sea Y una variable aleatoria binomial con $n = 1$ y probabilidad p de éxito.

a Encuentre la función de probabilidad y distribución para Y.

b Compare la función de distribución del inciso a con la del Ejercicio 4.3(a). ¿Qué concluye?

4.5 Suponga que Y es una variable aleatoria que toma sólo valores enteros 1, 2, . . . y tiene función de distribución $F(y)$. Demuestre que la función de probabilidad $p(y) = P(Y = y)$ está dada por

$$p(y) = \begin{cases} F(1), & y = 1, \\ F(y) - F(y-1), & y = 2, 3, \ldots \end{cases}$$

4.6 Considere una variable aleatoria con una distribución geométrica (Sección 3.5); esto es,

$$p(y) = q^{y-1}p, \qquad y = 1, 2, 3, \ldots, \quad 0 < p < 1.$$

a Demuestre que Y tiene función de distribución $F(y)$ tal que $F(i) = 1 - q^i$, $i = 0, 1, 2, \ldots$ y que, en general,

$$F(y) = \begin{cases} 0, & y < 0, \\ 1 - q^i, & i \leq y < i + 1, \qquad \text{para } i = 0, 1, 2, \ldots \end{cases}$$

b Demuestre que la función de distribución acumulativa precedente tiene las propiedades dadas en el Teorema 4.1.

4.7 Sea Y una variable aleatoria binomial con $n = 10$ y $p = .2$.

a Use la Tabla 1, Apéndice 3, para obtener $P(2 < Y < 5)$ y $P(2 \leq Y < 5)$. ¿Son iguales las probabilidades de que Y caiga en los intervalos (2, 5) y [2, 5)? ¿Por qué sí o por qué no?

b Use la Tabla 1, Apéndice 3, para obtener $P(2 < Y \le 5)$ y $P(2 \le Y \le 5)$. ¿Son iguales estas probabilidades? ¿Por qué sí o por qué no?

c Ya antes en esta sección dijimos que si Y es continua y $a < b$, entonces $P(a < Y < b) = P(a \le Y < b)$. ¿El resultado del inciso a contradice esta afirmación? ¿Por qué?

4.8 Suponga que Y tiene función de densidad

$$f(y) = \begin{cases} ky(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de k que haga de $f(y)$ una función de densidad de probabilidad.

b Encuentre $P(.4 \le Y \le 1)$.

c Encuentre $P(.4 \le Y < 1)$.

d Encuentre $P(Y \le .4 | Y \le .8)$.

e Encuentre $P(Y < .4 | Y < .8)$.

4.9 Una variable aleatoria Y tiene la siguiente función de distribución:

$$F(y) = P(Y \le y) = \begin{cases} 0, & \text{para } y < 2, \\ 1/8, & \text{para } 2 \le y < 2.5, \\ 3/16, & \text{para } 2.5 \le y < 4, \\ 1/2 & \text{para } 4 \le y < 5.5, \\ 5/8, & \text{para } 5.5 \le y < 6, \\ 11/16, & \text{para } 6 \le y < 7, \\ 1, & \text{para } y \ge 7. \end{cases}$$

a ¿Es Y una variable aleatoria continua o discreta? ¿Por qué?

b ¿Qué valores de Y son probabilidades positivas asignadas?

c Encuentre la función de probabilidad para Y.

d ¿Cuál es la mediana, $\phi_{.5}$, de Y?

4.10 Consulte la función de densidad dada en el Ejercicio 4.8.

a Encuentre el cuantil .95, $\phi_{.95}$, tal que $P(Y \le \phi_{.95}) = .95$.

b Encuentre un valor y_0 de manera que $P(Y < y_0) = .95$.

c Compare los valores para $\phi_{.95}$ y y_0 que obtuvo en los incisos a y b. Explique la relación entre estos dos valores.

4.11 Suponga que Y posee la función de densidad

$$f(y) = \begin{cases} cy, & 0 \le y \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de c que haga de $f(y)$ una función de densidad de probabilidad.

b Encuentre $F(y)$.

c Grafique $f(y)$ y $F(y)$.

d Use $F(y)$ para hallar $P(1 \le Y \le 2)$.

e Use $f(y)$ y geometría para hallar $P(1 \le Y \le 2)$.

4.12 El tiempo de falla (en cientos de horas) para un transistor es una variable aleatoria Y con función de distribución dada por

$$F(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-y^2}, & y \ge 0. \end{cases}$$

a Demuestre que $F(y)$ tiene las propiedades de una función de distribución.

b Encuentre el cuantil .30, $\phi_{.30}$, de Y.

c Encuentre $f(y)$.

d Encuentre la probabilidad de que el transistor opere durante al menos 200 horas.

e Encuentre $P(Y > 100 | Y \le 200)$.

4.13 Un proveedor de queroseno tiene un tanque de 150 galones que se llena al empezar cada semana. Su demanda semanal muestra un comportamiento de frecuencia relativo que aumenta de manera continua hasta 100 galones y luego se nivela entre 100 y 150 galones. Si Y denota la demanda semanal en cientos de galones, la frecuencia relativa de demanda puede ser modelada por

$$f(y) = \begin{cases} y, & 0 \le y \le 1, \\ 1, & 1 < y \le 1.5, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $F(y)$.

b Encuentre $P(0 \le Y \le .5)$.

c Encuentre $P(.5 \le Y \le 1.2)$.

4.14 Una gasolinera opera dos bombas, cada una de las cuales puede bombear hasta 10,000 galones de gasolina en un mes. La cantidad total de gasolina bombeada en un mes es una variable aleatoria Y (medida en 10,000 galones) con una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} y, & 0 < y < 1, \\ 2 - y, & 1 \le y < 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Grafique $f(y)$.

b Encuentre $F(y)$ y grafíquela.

c Encuentre la probabilidad de que la gasolinera bombee entre 8000 y 12,000 galones en un mes particular.

d Dado que la gasolinera bombeó más de 10,000 galones en un mes particular, encuentre la probabilidad de que haya bombeado más de 15,000 galones durante el mes.

4.15 Como una medición de inteligencia, a unos ratones se les toma el tiempo que tardan para pasar por un laberinto para llegar a una recompensa de alimento. El tiempo (en segundos) necesario para cualquier ratón es una variable aleatoria Y con una función de densidad dada por

$$f(y) = \begin{cases} \dfrac{b}{y^2}, & y \ge b, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde b es el tiempo mínimo posible necesario para recorrer el laberinto.

a Demuestre que $f(y)$ tiene las propiedades de una función de densidad.

b Encuentre $F(y)$.

c Encuentre $P(Y > b + c)$ para una constante positiva c.

d Si c y d son constantes positivas tales que $d > c$, encuentre $P(Y > b + d | Y > b + c)$.

4.16 Sea Y poseedor de una función de densidad

$$f(y) = \begin{cases} c(2 - y), & 0 \le y \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre c.

b Encuentre $F(y)$.

c Grafique $f(y)$ y $F(y)$.

d Use $F(y)$ en el inciso b para hallar $P(1 \le Y \le 2)$.

e Use geometría y la gráfica de $f(y)$ para calcular $P(1 \le Y \le 2)$.

4.17 El tiempo necesario para que estudiantes completen un examen de una hora es una variable aleatoria con una función de densidad dada por

$$f(y) = \begin{cases} cy^2 + y, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre c.

b Encuentre $F(y)$.

c Grafique $f(y)$ y $F(y)$.

d Use $F(y)$ que obtuvo en el inciso b para hallar $F(-1)$, $F(0)$ y $F(1)$.

e Encuentre la probabilidad de que un estudiante seleccionado al azar termine en menos de media hora.

f Dado que una estudiante particular necesita al menos 15 minutos para completar el examen, encuentre la probabilidad de que requiera al menos 30 minutos para terminar.

4.18 Tenga Y la función de densidad dada por

$$f(y) = \begin{cases} .2, & -1 < y \le 0, \\ .2 + cy, & 0 < y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre c.

b Encuentre $F(y)$.

c Grafique $f(y)$ y $F(y)$.

d Use $F(y)$ que obtuvo en el inciso b para hallar $F(-1)$, $F(0)$ y $F(1)$.

e Encuentre $P(0 \le Y \le .5)$.

f Encuentre $P(Y > .5 | Y > .1)$.

4.19 Sea la función de distribución de una variable aleatoria Y

$$F(y) = \begin{cases} 0, & y \le 0, \\ \dfrac{y}{8}, & 0 < y < 2, \\ \dfrac{y^2}{16}, & 2 \le y < 4, \\ 1, & y \ge 4. \end{cases}$$

a Encuentre la función de densidad de Y.

b Encuentre $P(1 \le Y \le 3)$.

c Encuentre $P(Y \ge 1.5)$.

d Encuentre $P(Y \ge 1 | Y \le 3)$.

4.3 Valores esperados para variables aleatorias continuas

El siguiente paso en el estudio de variables aleatorias continuas es hallar sus medias, varianzas y desviaciones estándar, con lo cual se adquieren medidas descriptivas numéricas asociadas con sus distribuciones. Muchas veces es difícil hallar la distribución de probabilidad para una variable aleatoria Y o una función de una variable aleatoria, $g(Y)$. Incluso si se conoce la función de densidad para una variable aleatoria, puede ser difícil evaluar integrales apropiadas (veremos que este es el caso cuando una variable aleatoria tiene una distribución gamma, Sección 4.6). Cuando encontramos estas situaciones, el comportamiento aproximado de variables de interés se puede establecer con el uso de sus momentos y la regla empírica o el teorema de Tchebysheff (Capítulos 1 y 3).

DEFINICIÓN 4.5 El valor esperado de una variable aleatoria continua Y es

$$E(Y) = \int_{-\infty}^{\infty} yf(y)\, dy,$$

siempre que exista la integral.[3]

Si la definición del valor esperado para una variable aleatoria discreta y, $E(Y) = \sum_y yp(y)$, es significativa, entonces la Definición 4.4 también debe estar de acuerdo con nuestra noción intuitiva de una media. La cantidad $f(y)dy$ corresponde a $p(y)$ para el caso discreto y la integración evoluciona de una sumatoria y es análoga a ella. En consecuencia, $E(Y)$ en la Definición 4.5 concuerda con nuestra noción de un promedio o media.

Al igual que en el caso discreto, en ocasiones estamos interesados en el valor esperado de una función de una variable aleatoria. Un resultado que nos permite evaluar ese valor esperado se da en el siguiente teorema.

TEOREMA 4.4 Sea $g(Y)$ una función de Y; entonces el valor esperado de $g(Y)$ está dado por

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)\, dy,$$

siempre que exista la integral.

La prueba del Teorema 4.4 es similar a la del Teorema 3.2 y se omite. Los valores esperados de tres importantes funciones de una variable aleatoria Y continua evolucionan como con-

3. Técnicamente, se dice que $E(Y)$ existe si

$$\int_{-\infty}^{\infty} |y| f(y)\, dy < \infty .$$

Este será el caso en todos los valores esperados que estudiemos y no mencionaremos esta condición adicional cada vez que definamos un valor esperado.

secuencia de teoremas de integración bien conocidos. Como es de esperarse, estos resultados llevan a conclusiones análogas a las contenidas en los Teoremas 3.3, 3.4 y 3.5. Por tanto, la prueba del Teorema 4.5 se dejará como ejercicio.

TEOREMA **4.5**

Sea c una constante y sean $g(Y)$, $g_1(Y)$, $g_2(Y)$, . . . , $g_k(Y)$ funciones de una variable aleatoria continua Y. Entonces se cumplen los siguientes resultados:

1. $E(c) = c$.
2. $E[cg(Y)] = cE[g(Y)]$.
3. $E[g_1(Y)+g_2(Y)+\cdots+g_k(Y)] = E[g_1(Y)]+E[g_2(Y)]+\cdots+E[g_k(Y)]$.

Al igual que en el caso de variables aleatorias discretas, frecuentemente buscamos el valor esperado de la función $g(Y) = (Y - \mu)^2$. Como antes, el valor esperado de esta función es la varianza de la variable aleatoria Y. Esto es, como en la Definición 3.5, $V(Y) = E(Y - \mu)^2$. Es un ejercicio sencillo demostrar que el Teorema 4.5 implica que $V(Y) = E(Y)^2 - \mu^2$.

EJEMPLO **4.6** En el Ejemplo 4.4 determinamos que $f(y) = (3/8)y^2$ para $0 \le y \le 2$, $f(y) = 0$ en cualquier otro punto, es una función de densidad válida. Si la variable aleatoria Y tiene esta función de densidad, encuentre $\mu = E(Y)$ y $\sigma^2 = V(Y)$.

Solución De acuerdo con la Definición 4.5,

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} yf(y)\,dy \\ &= \int_0^2 y\left(\frac{3}{8}\right)y^2\,dy \\ &= \left(\frac{3}{8}\right)\left(\frac{1}{4}\right)y^4\Big]_0^2 = 1.5. \end{aligned}$$

La varianza de Y se puede hallar una vez determinada $E(Y^2)$. En este caso,

$$\begin{aligned} E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y)\,dy \\ &= \int_0^2 y^2\left(\frac{3}{8}\right)y^2\,dy \\ &= \left(\frac{3}{8}\right)\left(\frac{1}{5}\right)y^5\Big]_0^2 = 2.4. \end{aligned}$$

Por tanto, $\sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2 = 2.4 - (1.5)^2 = 0.15$. ■

Ejercicios

4.20 Si, como en el Ejercicio 4.16, Y tiene función de densidad

$$f(y) = \begin{cases} (1/2)(2-y), & 0 \le y \le 2, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

encuentre la media y la varianza de Y.

4.21 Si, como en el Ejercicio 4.17, Y tiene función de densidad

$$f(y) = \begin{cases} (3/2)y^2 + y, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

encuentre la media y la varianza de Y.

4.22 Si, como en el Ejercicio 4.18, Y tiene función de densidad

$$f(y) = \begin{cases} .2, & -1 < y \le 0, \\ .2 + (1.2)y, & 0 < y \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

encuentre la media y la varianza de Y.

4.23 Demuestre el Teorema 4.5.

4.24 Si Y es una variable aleatoria continua con función de densidad $f(y)$, use el Teorema 4.5 para demostrar que $\sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2$.

4.25 Si, como en el Ejercicio 4.19, Y tiene función de distribución

$$F(y) = \begin{cases} 0, & y \le 0, \\ \dfrac{y}{8}, & 0 < y < 2, \\ \dfrac{y^2}{16}, & 2 \le y < 4, \\ 1, & y \ge 4, \end{cases}$$

encuentre la media y la varianza de Y.

4.26 Si Y es una variable aleatoria continua con media μ y varianza σ^2 y a y b son constantes, use el Teorema 4.5 para demostrar lo siguiente:

a $E(aY + b) = aE(Y) + b = a\mu + b$.

b $V(aY + b) = a^2V(Y) = a^2\sigma^2$.

4.27 Para ciertas muestras de minerales, la proporción Y de impurezas por muestra es una variable aleatoria con función de densidad dada en el Ejercicio 4.21. El valor en dólares de cada muestra es $W = 5 - .5Y$. Encuentre la media y la varianza de W.

4.28 La proporción de tiempo por día en la que todas las cajas de un supermercado están ocupadas, es una variable aleatoria Y con función de densidad

$$f(y) = \begin{cases} cy^2(1-y)^4, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de c que haga de $f(y)$ una función de densidad de probabilidad.

b Encuentre $E(Y)$.

4.29 La temperatura Y a la que se conecta un interruptor controlado por un termostato tiene función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 1/2, & 59 \leq y \leq 61, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $E(Y)$ y $V(Y)$.

4.30 La proporción de tiempo Y en la que un robot industrial está en operación durante una semana de 40 horas es una variable aleatoria con función de densidad de probabilidad

$$f(y) = \begin{cases} 2y, & 0 \leq y \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $E(Y)$ y $V(Y)$.

b Para el robot motivo de estudio, la utilidad X para una semana está dada por $X = 200Y - 60$. Encuentre $E(X)$ y $V(X)$.

c Encuentre un intervalo en el que la utilidad sea de al menos 75% durante las semanas que el robot esté en uso.

4.31 La radiación solar total diaria para un lugar específico en Florida durante octubre tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} (3/32)(y - 2)(6 - y), & 2 \leq y \leq 6, \\ 0, & \text{en otro lugar.} \end{cases}$$

con mediciones en cientos de calorías. Encuentre la radiación solar diaria esperada para octubre.

4.32 El tiempo semanal de un CPU empleado por una firma de contadores tiene función de densidad de probabilidad (medida en horas) dada por

$$f(y) = \begin{cases} (3/64)y^2(4 - y), & 0 \leq y \leq 4, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor esperado y la varianza de tiempo semanal del CPU.

b El tiempo del CPU cuesta \$200 por hora a la empresa. Encuentre el valor esperado y la varianza del costo semanal para el CPU.

c ¿Esperaría usted que el costo semanal esperado exceda de \$600 con frecuencia? ¿Por qué?

4.33 El pH de muestras de agua para un lago específico es una variable aleatoria Y con función de densidad de probabilidad dada por

$$f(y) = \begin{cases} (3/8)(7 - y)^2, & 5 \leq y \leq 7, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $E(Y)$ y $V(Y)$.

b Encuentre un intervalo más corto que (5, 7) en el que deban estar al menos tres cuartos de las mediciones de pH.

c ¿Esperaría ver con frecuencia una medición de pH debajo de 5.5? ¿Por qué?

***4.34** Suponga que Y es una variable aleatoria continua con densidad $f(y)$ que es positiva sólo si $y \geq 0$. Si $F(y)$ es la función de distribución, demuestre que

$$E(Y) = \int_0^\infty y f(y)\, dy = \int_0^\infty [1 - F(y)]\, dy.$$

[*Sugerencia:* Si $y > 0$, $y = \int_0^y dt$, y $E(Y) = \int_0^\infty y f(y)\, dy = \int_0^\infty \left\{\int_0^y dt\right\} f(y)\, dy$. Intercambie el orden de integración para obtener el resultado deseado.][4]

4. Los ejercicios precedidos por un asterisco son opcionales.

*4.35 Si Y es una variable aleatoria continua tal que $E[(Y-a)^2] < \infty$ para toda a, demuestre que $E[(Y-a)^2]$ se minimiza cuando $a = E(Y)$. [*Sugerencia:* $E[(Y-a)^2] = E(\{[Y - E(Y)] + [E(Y) - a]\}^2)$.]

*4.36 ¿El resultado obtenido en el Ejercicio 4.35 también es válido para variables aleatorias discretas? ¿Por qué?

*4.37 Si Y es una variable aleatoria continua con función de densidad $f(y)$ que es simétrica alrededor de 0 (es decir, $f(y) = f(-y)$ para toda y) y $E(Y)$ existe, demuestre que $E(Y) = 0$. [*Sugerencia:* $E(Y) = \int_{-\infty}^{0} yf(y)\,dy + \int_{0}^{\infty} yf(y)\,dy$. Haga el cambio de variable $w = -y$ en la primera integral.]

4.4 La distribución de probabilidad uniforme

Suponga que un autobús llega siempre a una parada particular entre las 8:00 y las 8:10 a.m. y que la probabilidad de que llegue en cualquier subintervalo dado es proporcional sólo a la duración del subintervalo. Esto es, es igual de probable que llegue entre las 8:00 y 8:02 a que llegue entre las 8:06 y las 8:08. Denote con Y el tiempo que una persona deba esperar para que llegue el autobús si llegó a la parada exactamente a las 8:00. Si con cuidado medimos en minutos cuánto tiempo después de las 8:00 llegó el autobús en varias mañanas, podríamos desarrollar un histograma de frecuencia relativa para los datos.

A partir de la descripción que acabamos de dar, debe ser evidente que la frecuencia relativa con la cual observamos un valor de Y entre 0 y 2 sería aproximadamente la misma que la frecuencia relativa con la cual observamos un valor de Y entre 6 y 8. Un modelo razonable para la función de densidad de Y se muestra en la Figura 4.9. Como las áreas bajo las curvas representan probabilidades para variables aleatorias continuas y $A_1 = A_2$ (por inspección), se deduce que $P(0 \le Y \le 2) = P(6 \le Y \le 8)$, como se desea.

La variable aleatoria Y que acabamos de examinar es un ejemplo de una variable aleatoria que tiene una distribución uniforme. La forma general para la función de densidad de una variable aleatoria con una distribución uniforme es como sigue.

DEFINICIÓN 4.6 Si $\theta_1 < \theta_2$, se dice que una variable aleatoria Y tiene *distribución de probabilidad uniforme* en el intervalo (θ_1, θ_2) si y sólo si la función de densidad de Y es

$$f(y) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & \theta_1 \le y \le \theta_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

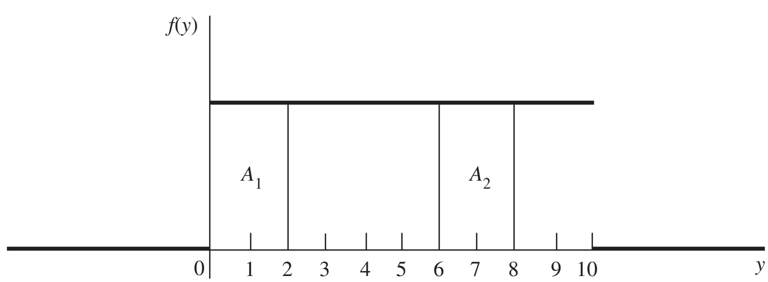


FIGURA **4.9** Función de densidad para Y

En el problema del autobús podemos tomar $\theta_1 = 0$ y $\theta_2 = 10$ porque estamos interesados sólo en un intervalo particular de diez minutos. La función de densidad que se estudia en el Ejemplo 4.2 es una distribución uniforme con $\theta_1 = 0$ y $\theta_2 = 1$. Las gráficas de la función de distribución y función de densidad para la variable aleatoria del Ejemplo 4.2 se dan en las Figuras 4.4 y 4.5, respectivamente.

DEFINICIÓN **4.7**

Las constantes que determinan la forma específica de una función de densidad se denominan *parámetros* de la función de densidad.

Las cantidades θ_1 y θ_2 son parámetros de la función de densidad uniforme y son valores numéricos claramente significativos asociados con la función de densidad teórica. Tanto la amplitud como la probabilidad de que Y caiga en cualquier intervalo determinado dependen de los valores de θ_1 y θ_2.

Algunas variables aleatorias continuas en física, administración y ciencias biológicas tienen distribuciones de probabilidad aproximadamente uniformes. Por ejemplo, suponga que el número de eventos, como las llamadas que entran en un conmutador, que se presentan en el intervalo $(0, t)$ tienen una distribución de Poisson. Si se sabe que exactamente uno de estos eventos ha ocurrido en el intervalo $(0, t)$, entonces el tiempo real del suceso está distribuido de manera uniforme en este intervalo.

EJEMPLO **4.7** La llegada de clientes a una caja en un establecimiento sigue una distribución de Poisson. Se sabe que durante un periodo determinado de 30 minutos, un cliente llega a la caja. Encuentre la probabilidad de que el cliente llegue durante los últimos 5 minutos del periodo de 30 minutos.

Solución Como acabamos de citar, el tiempo real de llegada sigue una distribución uniforme en el intervalo de (0, 30). Si Y denota el tiempo de llegada, entonces

$$P(25 \leq Y \leq 30) = \int_{25}^{30} \frac{1}{30}\, dy = \frac{30 - 25}{30} = \frac{5}{30} = \frac{1}{6}.$$

La probabilidad de que la llegada ocurra en cualquier otro intervalo de 5 minutos también es 1/6. ■

Como veremos, la distribución uniforme es muy importante por razones teóricas. Los estudios de simulación son técnicas valiosas para validar modelos en estadística. Si deseamos un conjunto de observaciones de una variable aleatoria Y con función de distribución $F(y)$, a menudo podemos obtener los resultados deseados si transformamos un conjunto de observaciones en una variable aleatoria uniforme. Por esta razón, casi todos los sistemas de cómputo contienen un generador de números aleatorios que produce valores observados para una variable aleatoria que tiene una distribución uniforme continua.

TEOREMA 4.6 Si $\theta_1 < \theta_2$ y Y es una variable aleatoria uniformemente distribuida en el intervalo (θ_1, θ_2), entonces

$$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2} \quad \text{y} \quad \sigma^2 = V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Prueba Por la Definición 4.5,

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f(y)\, dy \\ &= \int_{\theta_1}^{\theta_2} y \left(\frac{1}{\theta_2 - \theta_1} \right) dy \\ &= \left(\frac{1}{\theta_2 - \theta_1} \right) \frac{y^2}{2} \bigg]_{\theta_1}^{\theta_2} = \frac{\theta_2^2 - \theta_1^2}{2(\theta_2 - \theta_1)} \\ &= \frac{\theta_2 + \theta_1}{2}. \end{aligned}$$

Observe que la media de una variable aleatoria uniforme es simplemente el valor que está a la mitad entre los valores de los dos parámetros, θ_1 y θ_2. La obtención de la varianza se deja como ejercicio.

Ejercicios

4.38 Suponga que Y tiene una distribución uniforme en el intervalo $(0, 1)$.

a Encuentre $F(y)$.

b Demuestre que $P(a \leq Y \leq a + b)$, para $a \geq 0$, $b \geq 0$, y $a + b \leq 1$ depende sólo del valor de b.

4.39 Si una paracaidista aterriza en un punto aleatorio en una recta entre los marcadores A y B, encuentre la probabilidad de que ella esté más cerca de A que de B. Encuentre la probabilidad de que su distancia hasta A sea más de tres veces su distancia a B.

4.40 Suponga que tres paracaidistas operan de manera independiente como se describe en el Ejercicio 4.39. ¿Cuál es la probabilidad de que exactamente uno de los tres aterrice en el punto medio entre A y B?

4.41 Una variable aleatoria Y tiene una distribución uniforme en el intervalo (θ_1, θ_2). Obtenga la varianza de Y.

4.42 La *mediana* de la distribución de una variable aleatoria continua Y es el valor $\phi_{.5}$ de manera que $P(Y \leq \phi_{.5}) = 0.5$. ¿Cuál es la mediana de la distribución uniforme en el intervalo (θ_1, θ_2)?

4.43 Un círculo de radio r tiene área $A = \pi r^2$. Si un círculo aleatorio tiene un radio que está uniformemente distribuido en el intervalo $(0, 1)$, ¿cuáles son la media y la varianza del área del círculo?

4.44 El cambio en profundidad de un río de un día al siguiente, medida (en pies) en un lugar específico, es una variable aleatoria Y con la siguiente función de densidad:

$$f(y) = \begin{cases} k, & -2 \leq y \leq 2 \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Determine el valor de k.

b Obtenga la función de distribución para Y.

4.45 Al estudiar bajas cotizaciones para contratos de embarques, una empresa fabricante de microcomputadoras encuentra que los contratos interestatales tienen bajas cotizaciones que están uniformemente distribuidas entre 20 y 25, en unidades de miles de dólares. Encuentre la probabilidad de que la baja cotización en el siguiente contrato interestatal

a esté por debajo de \$22,000.

b sea de más de \$24,000.

4.46 Consulte el Ejercicio 4.45. Encuentre el valor esperado de bajas cotizaciones en contratos del tipo descrito ahí.

4.47 La falla de una tarjeta de circuito que utiliza un sistema de cómputo interrumpe el trabajo hasta que se instala una nueva. El tiempo de entrega, Y, está uniformemente distribuido en el intervalo de uno a cinco días. El costo de la falla de una tarjeta y la interrupción incluye el costo fijo c_0 de una nueva tarjeta y un costo que aumenta proporcionalmente con Y^2. Si C es el costo en que se incurre, $C = c_0 + c_1 Y^2$.

a Encuentre la probabilidad de que el tiempo de entrega exceda de dos días.

b En términos de c_0 y c_1, encuentre el costo esperado asociado con una sola tarjeta de circuito que falle.

4.48 Si un punto se localiza *al azar* en un intervalo (a, b) y si Y denota la ubicación del punto, entonces se supone que Y tiene una distribución uniforme en (a, b). Una experta en eficiencia de la planta selecciona al azar un lugar, a lo largo de una línea de ensamble de 500 pies, desde el cual observa hábitos de los trabajadores de la línea. ¿Cuál es la probabilidad de que el punto que ella seleccione se encuentre

a a no más de 25 pies del final de la línea?

b a no más de 25 pies del principio de la línea?

c más cerca del principio de la línea que al final de la línea?

4.49 Una llamada telefónica llega a un conmutador al azar en un intervalo de no más de un minuto. El conmutador estuvo totalmente ocupado durante 15 segundos en este periodo de un minuto. ¿Cuál es la probabilidad de que la llamada llegara cuando el conmutador no hubiera estado totalmente ocupado?

4.50 Empezando a las 12:00 de la noche, un centro de computadoras funciona durante una hora y deja de operar dos horas en un ciclo regular. Una persona que desconoce este horario marca al centro en una hora al azar entre las 12:00 de la noche y las 5:00 a.m. ¿Cuál es la probabilidad de que el centro esté funcionando cuando entre la llamada de la persona?

4.51 El tiempo de ciclo para camiones que transportan concreto al lugar de construcción de una carretera está uniformemente distribuido en el intervalo de 50 a 70 minutos. ¿Cuál es la probabilidad de que el tiempo de ciclo exceda de 65 minutos si se sabe que el tiempo de ciclo excede de 55 minutos?

4.52 Consulte el Ejercicio 4.51. Encuentre la media y la varianza de los tiempos de ciclo para los camiones.

4.53 El número de tarjetas de circuito defectuosas que salen de una máquina soldadora sigue una distribución de Poisson. Durante un día específico de ocho horas, se encontró una tarjeta defectuosa.

a Encuentre la probabilidad de que haya sido producida durante la primera hora de operación durante ese día.

b Encuentre la probabilidad de que haya sido producida durante la última hora de operación durante ese día.

c Dado que no se produjeron tarjetas defectuosas durante las primeras cuatro horas de operación, encuentre la probabilidad de que la tarjeta defectuosa se fabricara durante la quinta hora.

4.54 Al usar el método de triangulación para determinar el alcance de una sonda acústica, el equipo de prueba debe medir con precisión el tiempo que tarda en llegar el frente de onda esférica a un sensor de recepción.

De acuerdo con Perruzzi y Hilliard (1984), los errores de medición se pueden modelar como si tuvieran una distribución uniforme de –0.05 a +0.05 μs (microsegundos).

a ¿Cuál es la probabilidad de que una medición de tiempo de llegada sea precisa con tolerancia de 0.01 μs?

b Encuentre la media y varianza de los errores de medición.

4.55 Consulte el Ejercicio 4.54. Suponga que los errores de medición están uniformemente distribuidos entre –0.02 a +0.05 μs.

a ¿Cuál es la probabilidad de que una medición particular de tiempo de llegada sea precisa con tolerancia de no más de 0.01 μs?

b Encuentre la media y varianza de los errores de medición.

4.56 Consulte el Ejemplo 4.7. Encuentre la probabilidad condicional de que un cliente llegue durante los últimos 5 minutos del periodo de 30 minutos, si se sabe que ninguno llega durante los primeros 10 minutos del periodo.

4.57 De acuerdo con Zimmels (1983), los tamaños de partículas empleadas en experimentos de sedimentación a menudo tienen una distribución uniforme. En sedimentación que comprenda mezclas de partículas de varios tamaños, las más grandes impiden los movimientos de las más pequeñas. Entonces, es importante estudiar la media y la varianza de los tamaños de partículas. Suponga que las partículas esféricas tienen diámetros que están uniformemente distribuidos entre .01 y .05 centímetros. Encuentre la media y la varianza de los *volúmenes* de estas partículas. (Recuerde que el volumen de una esfera es $(4/3)\pi r^3$.)

4.5 La distribución de probabilidad normal

La distribución de probabilidad continua que más se utiliza es la distribución normal, con la conocida forma de campana que estudiamos en relación con la regla empírica. Los ejemplos y ejercicios de esta sección ilustran algunas de las numerosas variables aleatorias que tienen distribuciones que se calculan en forma muy cercana por medio de una distribución de probabilidad normal. En el Capítulo 7 presentaremos un argumento que explica, al menos parcialmente, el suceso común de distribuciones normales de datos en la naturaleza. La función de densidad normal es como sigue:

DEFINICIÓN **4.8**

Se dice que una variable Y tiene una *distribución normal de probabilidad* si y sólo si, para $\sigma > 0$ y $-\infty < \mu < \infty$, la función de densidad de Y es

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}, \qquad -\infty < y < \infty.$$

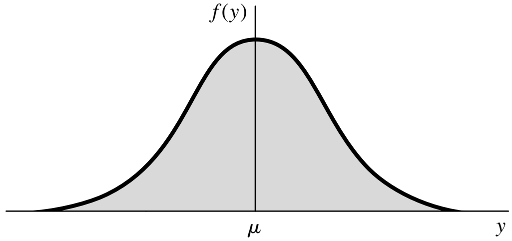
Observe que la función de densidad normal contiene dos parámetros, μ y σ.

TEOREMA **4.7**

Si Y es una variable aleatoria normalmente distribuida con parámetros μ y σ, entonces

$$E(Y) = \mu \quad \text{y} \quad V(Y) = \sigma^2.$$

FIGURA **4.10** La función de densidad de probabilidad normal



La demostración de este teorema se difiere a la Sección 4.9, donde obtendremos la función generadora de momento de una variable aleatoria normalmente distribuida. Los resultados contenidos en el Teorema 4.7 implican que el parámetro μ localiza el centro de la distribución y que σ mide su dispersión. Una gráfica de una función de densidad normal se muestra en la Figura 4.10.

Las áreas bajo la función de densidad normal correspondientes a $P(a \leq Y \leq b)$ requieren la evaluación de la integral

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}\, dy.$$

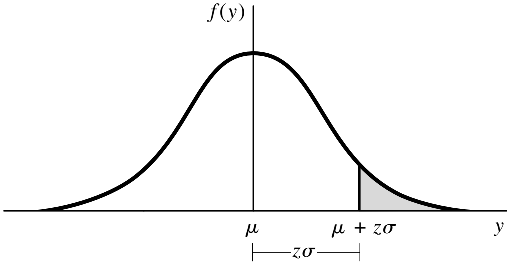
Desafortunadamente, no existe una expresión de forma cerrada para esta integral; en consecuencia, su evaluación requiere el uso de técnicas de integración numérica. Las probabilidades y cuantiles para variables aleatorias con distribuciones normales se encuentran fácilmente usando *R* y *S*–Plus. Si Y tiene una distribución normal con media μ y desviación estándar σ, el comando `pnorm(`y_0`,`μ`,`σ`)` de *R* (o *S*–Plus) genera $P(Y \leq y_0)$ mientras que `qnorm(p,`μ`,`σ`)` da el p–ésimo cuantil, el valor de ϕ_p tal que $P(Y \leq \phi_p) = p$. Aun cuando hay un número infinito de distribuciones normales (μ puede tomar cualquier valor finito, en tanto que σ puede tomar cualquier valor finito positivo), sólo necesitamos una tabla —la Tabla 4, Apéndice 3— para calcular áreas bajo densidades normales. Las probabilidades y cuantiles asociados con variables aleatorias normalmente distribuidas también se pueden hallar usando la aplicación breve (applet) *Normal Tail Areas and Quantiles* accesibles en www.thomsonedu.com/statistics/wackerly. El único beneficio real obtenido al usar software para obtener probabilidades y cuantiles asociados con variables aleatorias normalmente distribuidas, es que el software da respuestas que son correctas hasta un gran número de lugares decimales.

La función de densidad normal es simétrica alrededor del valor μ, de modo que las áreas tienen que ser tabuladas en sólo un lado de la media. Las áreas tabuladas están a la derecha de los puntos z, donde z es la distancia desde la media, medida en desviaciones estándar. Esta área está sombreada en la Figura 4.11.

EJEMPLO **4.8** Denote con Z una variable aleatoria normal con media 0 y desviación estándar 1.

a Encuentre $P(Z > 2)$.
b Encuentre $P(-2 \leq Z \leq 2)$.
c Encuentre $P(0 \leq Z \leq 1.73)$.

FIGURA **4.11**
Área tabulada para la función de densidad normal



Solución **a** Como $\mu = 0$ y $\sigma = 1$, el valor 2 está en realidad a $z = 2$ desviaciones estándar arriba de la media. Avance hacia abajo en la primera columna (z) de la Tabla 4, Apéndice 3, y lea el área opuesta a $z = 2.0$. Esta área, denotada por el símbolo $A(z)$, es $A(2.0) = .0228$. Entonces, $P(Z > 2) = .0228$.

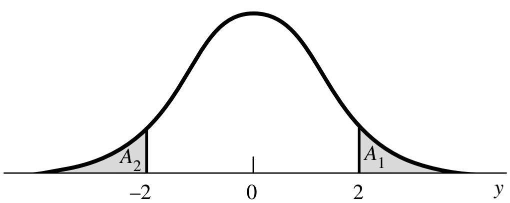
b Consulte la Figura 4.12, donde hemos sombreado el área de interés. En el inciso a determinamos que $A_1 = A(2.0) = 0.228$. Como la función de densidad es simétrica alrededor de la media $\mu = 0$, se deduce que $A_2 = A_1 = .0228$ y por tanto que

$$P(-2 \leq Z \leq 2) = 1 - A_1 - A_2 = 1 - 2(.0228) = .9544.$$

c Como $P(Z > 0) = A(0) = .5$, obtenemos que $P(0 \leq Z \leq 1.73) = .5 - A(1.73)$, donde $A(1.73)$ se obtiene al bajar por la columna z de la Tabla 4, Apéndice 3, a la entrada 1.7 y luego en sentido horizontal por la parte superior de la tabla a la columna marcada .03 para leer $A(1.73) = .0418$. De esta manera,

$$P(0 \leq Z \leq 1.73) = .5 - .0418 = .4582.$$

FIGURA **4.12**
Área deseada para el Ejemplo 4.8(b)



■

EJEMPLO **4.9** Las calificaciones para un examen de admisión a una universidad están normalmente distribuidas con media de 75 y desviación estándar 10. ¿Qué fracción de las calificaciones se encuentra entre 80 y 90?

Solución Recuerde que z es la distancia desde la media de una distribución normal expresada en unidades de desviación estándar. Entonces,

$$z = \frac{y - \mu}{\sigma}.$$

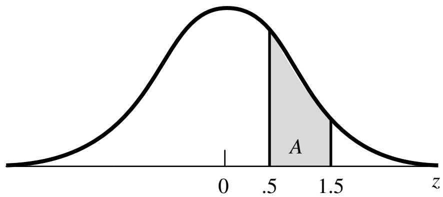


FIGURA **4.13**
Área requerida para el Ejemplo 4.9

Entonces la fracción deseada de la población está dada por el área entre

$$z_1 = \frac{80 - 75}{10} = .5 \quad \text{y} \quad z_2 = \frac{90 - 75}{10} = 1.5.$$

Esta área está sombreada en la Figura 4.13.

Usted puede ver en la Figura 4.13 que $A = A(.5) - A(1.5) = .3085 - .0668 = .2417$. ■

Siempre podemos transformar una variable aleatoria normal Y en una variable aleatoria *normal estándar* Z si usamos la relación

$$Z = \frac{Y - \mu}{\sigma}.$$

La Tabla 4, Apéndice 3 se puede usar para calcular probabilidades, como se muestra aquí. Z localiza un punto medido desde la media de una variable aleatoria normal, con la distancia *expresada en unidades de la desviación estándar* de la variable aleatoria normal original. Entonces, el valor medio de Z debe ser 0 y su desviación estándar debe ser igual a 1. La prueba de que la *variable aleatoria normal estándar,* Z, está normalmente distribuida con media 0 y desviación estándar 1 se proporciona en el Capítulo 6.

En la aplicación breve *Normal Probabilities*, accesible en www.thomsonedu.com/statistics/wackerly, se ilustra la correspondencia entre probabilidades normales en las escalas originales y transformadas (z). Para contestar la pregunta planteada en el Ejemplo 4.9, localice el intervalo de interés, (80, 90), en el eje horizontal inferior marcado como Y. Las calificaciones z correspondientes se dan en el eje horizontal superior y es evidente que el área sombreada da $P(80 < Y < 90) = P(0.5 < Z < 1.5) = 0.2417$ (vea la Figura 4.14). Algunos de los ejercicios del final de esta sección sugieren que el estudiante use esta aplicación para reforzar los cálculos de probabilidades asociadas con variables aleatorias normalmente distribuidas.

Ejercicios

4.58 Use la Tabla 4, Apéndice 3 para hallar las siguientes probabilidades para una variable Z aleatoria normal estándar:

a $P(0 \le Z \le 1.2)$.
b $P(-.9 \le Z \le 0)$.
c $P(.3 \le Z \le 1.56)$.

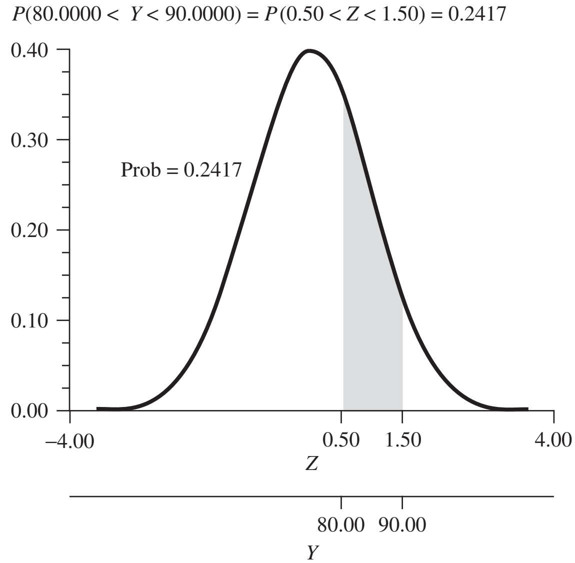


FIGURA **4.14** Área requerida para el Ejemplo 4.9, que usa escalas original y transformada (z).

d $P(-.2 \leq Z \leq .2)$.

e $P(-1.56 \leq Z \leq -.2)$.

f **Ejercicio Applet** Use la aplicación breve *Normal Probabilities* para obtener $P(0 \leq Z \leq 1.2)$. ¿Por qué son idénticos los valores dados en los dos ejes horizontales?

4.59 Si Z es una variable aleatoria normal estándar, encuentre el valor z_0 tal que

a $P(Z > z_0) = .5$.

b $P(Z < z_0) = .8643$.

c $P(-z_0 < Z < z_0) = .90$.

d $P(-z_0 < Z < z_0) = .99$.

4.60 Una variable aleatoria normalmente distribuida tiene función de densidad

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(y-\mu)^2/(2\sigma^2)}, \qquad -\infty < y < \infty.$$

Usando las propiedades fundamentales asociadas con cualquier función de densidad, demuestre que el parámetro σ debe ser tal que $\sigma > 0$.

4.61 ¿Cuál es la mediana de una variable aleatoria normalmente distribuida con media μ y desviación estándar σ?

4.62 Si Z es una variable aleatoria normal estándar, ¿cuál es

a $P(Z^2 < 1)$?

b $P(Z^2 < 3.84146)$?

4.63 Una compañía que manufactura y embotella jugo de manzana usa una máquina que automáticamente llena botellas de 16 onzas. Hay alguna variación, no obstante, en las cantidades de líquido que se ponen en las botellas que se llenan. Se ha observado que la cantidad de líquido está normalmente distribuida en forma aproximada con media de 16 onzas y desviación estándar de 1 onza.

a Use la Tabla 4, Apéndice 3 para determinar la proporción de botellas que tendrán más de 17 onzas.

b **Ejercicio Applet** Use la aplicación breve *Normal Probabilities* para obtener la respuesta al inciso a.

4.64 Se observó que la cantidad semanal de dinero gastado por una compañía durante largo tiempo en mantenimiento y reparaciones, está normalmente distribuida en forma aproximada con media de \$400 y desviación estándar de \$20. Si están presupuestados \$450 para la próxima semana, ¿cuál es la probabilidad de que los costos reales rebasen la cantidad presupuestada?

a Conteste la pregunta usando la Tabla 4, Apéndice 3.

b **Ejercicio Applet** Use la aplicación breve *Normal Probabilities* para obtener la respuesta.

c ¿Por qué son diferentes los valores marcados en los dos ejes horizontales?

4.65 En el Ejercicio 4.64, ¿cuánto debe presupuestarse para reparaciones y mantenimiento semanal para lograr que la probabilidad de que la cantidad presupuestada en una semana determinada sea excedida sólo .1?

4.66 Una operación de maquinado produce cojinetes con diámetros que están normalmente distribuidos con media de 3.0005 pulgadas y desviación estándar de .0010 pulgadas. Las especificaciones requieren que los diámetros de los cojinetes se encuentren en el intervalo $3.000 \pm .0020$ pulgadas. Los cojinetes que estén fuera de este intervalo son considerados de desecho y deben volver a maquinarse. Con el ajuste de la máquina existente, ¿qué fracción de la producción total se desechará?

a Conteste la pregunta usando la Tabla 4, Apéndice 3.

b **Ejercicio Applet** Obtenga la respuesta usando la aplicación breve *Normal Probabilities.*

4.67 En el Ejercicio 4.66, ¿cuál debe ser el diámetro medio para que la fracción de cojinetes desechados sea mínima?

4.68 Los promedios de calificaciones (GPA, por sus siglas en inglés) de una gran población de estudiantes universitarios están normalmente distribuidos en forma aproximada, con media de 2.4 y desviación estándar .8. ¿Qué fracción de los estudiantes alcanzarán un GPA de más de 3.0?

a Conteste la pregunta usando la Tabla 4, Apéndice 3.

b **Ejercicio Applet** Use la aplicación breve *Normal Tail Areas and Quantiles* para obtener la respuesta.

4.69 Consulte el Ejercicio 4.68. Si los estudiantes que alcancen un GPA menor que 1.9 serán suspendidos de la universidad, ¿qué porcentaje de los estudiantes será suspendido?

4.70 Consulte el Ejercicio 4.68. Suponga que se seleccionan al azar tres estudiantes del alumnado. ¿Cuál es la probabilidad de que los tres alcancen un GPA de más de 3.0?

4.71 Se especifica que los cables manufacturados para usarse en un sistema de computadora deben tener resistencias entre .12 y .14 ohms. Las resistencias medidas reales de los cables producidos por la compañía *A* tienen una distribución de probabilidad normal con media de .13 ohms y desviación estándar .005 ohm.

a ¿Cuál es la probabilidad de que un cable seleccionado al azar de la producción de la compañía *A* satisfaga las especificaciones?

b Si cuatro de estos cables se usan en el sistema de cada computadora y todos son seleccionados de la compañía *A*, ¿cuál es la probabilidad de que los cuatro en un sistema seleccionado al azar satisfagan las especificaciones?

4.72 Un método para llegar a pronósticos económicos es usar un método de consenso. Un pronóstico se obtiene de todos y cada uno de un gran número de analistas; el promedio de los pronósticos de estas personas es el pronóstico de consenso. Suponga que los pronósticos de tasa de interés preferencial de enero de 1996 de todos los analistas económicos están distribuidos normalmente en forma aproximada con media de 7% y desviación estándar de 2.6%. Si un solo analista se selecciona al azar entre este grupo, ¿cuál es la probabilidad de que el pronóstico del analista de la tasa de interés preferencial

a exceda de 11%?

b sea menor que 9%?

4.73 El ancho de rollos de tela está normalmente distribuido con media de 950 mm (milímetros) y desviación estándar de 10 mm.

a ¿Cuál es la probabilidad de que un rollo seleccionado al azar tenga un ancho de entre 947 y 958 mm?

b ¿Cuál es el valor apropiado para C de manera que un rollo seleccionado al azar tenga un ancho menor que C con probabilidad .8531?

4.74 Se supone que las calificaciones de un examen están normalmente distribuidas con media de 78 y varianza de 36.

a ¿Cuál es la probabilidad de que una persona que haga el examen alcance calificaciones mayores de 72?

b Suponga que los estudiantes que alcancen el 10% más alto de esta distribución reciben una calificación de A. ¿Cuál es la calificación mínima que un estudiante debe recibir para ganar una calificación de A?

c ¿Cuál debe ser el punto límite para pasar el examen si el examinador desea pasar sólo a 28.1% más alto de todas las calificaciones?

d ¿Aproximadamente qué proporción de estudiantes tienen calificaciones de 5 o más puntos arriba de la calificación que corta al 25% más bajo?

e **Ejercicio Applet** Conteste los incisos a–d usando la aplicación breve *Normal Tail Areas and Quantiles.*

f Si se sabe que la calificación de un estudiante excede de 72, ¿cuál es la probabilidad de que su calificación exceda de 84?

4.75 Una máquina expendedora de bebidas gaseosas puede ser regulada para descargar un promedio de μ onzas por vaso. Si las onzas están normalmente distribuidas con desviación estándar de 0.3 onzas, determine los valores para μ de modo que vasos de 8 onzas se sirvan sólo 1% del tiempo.

4.76 La máquina descrita en el Ejercicio 4.75 tiene desviación estándar σ que se puede fijar en ciertos niveles al ajustar la máquina con todo cuidado. ¿Cuál es el máximo valor de σ que permitirá que la cantidad real servida esté a no más de 1 onza de la media con probabilidad de al menos .95?

4.77 Los exámenes de admisión SAT y ACT (de aptitud y universitario) se aplican a miles de estudiantes cada año. Las secciones de matemáticas de cada uno de estos exámenes producen calificaciones que están normalmente distribuidas, en forma aproximada. En años recientes las calificaciones de exámenes SAT de matemáticas han promediado 480 con desviación estándar de 100. El promedio y desviación estándar para calificaciones ACT de matemáticas son 18 y 6, respectivamente.

a Una escuela de ingeniería establece 550 como calificación mínima SAT de matemáticas para estudiantes de nuevo ingreso. ¿Qué porcentaje de estudiantes obtendrá una calificación por debajo de 550 en un año típico?

b ¿Qué calificación debe establecer la escuela de ingeniería como estándar comparable en el examen ACT de matemáticas?

4.78 Demuestre que el máximo valor de la densidad normal con parámetros μ y σ es $1/(\sigma\sqrt{2\pi})$ y sucede cuando $y = \mu$.

4.79 Demuestre que la densidad normal con parámetros μ y σ tiene puntos de inflexión en los valores $\mu - \sigma$ y $\mu + \sigma$. (Recuerde que un punto de inflexión es aquel donde la curva cambia de dirección de cóncava hacia arriba a cóncava hacia abajo o viceversa y ocurre cuando la segunda derivada cambia de signo. Este cambio en signo puede presentarse cuando la segunda derivada es igual a cero.)

4.80 Suponga que Y está normalmente distribuida con media μ y desviación estándar σ. Después de observar el valor de Y, un matemático construye un rectángulo con longitud $L = |Y|$ y ancho $W = 3|Y|$. Denote con A el área del triángulo resultante. ¿Cuál es $E(A)$?

4.6 La distribución de probabilidad gamma

Algunas variables aleatorias son siempre no negativas y por varias razones dan distribuciones de datos que está sesgadas (no simétricas) a la derecha. Esto es, casi toda el área bajo la función de densidad está ubicada cerca del origen y la función de densidad cae gradualmente conforme y aumenta. En la Figura 4.15 se muestra una función de densidad de probabilidad sesgada.

Los intervalos de tiempo entre mal funcionamiento de motores de aviones poseen una distribución de frecuencia sesgada, al igual que los intervalos de llegada en una fila de espera en las cajas de un supermercado (esto es, la fila de espera para llegar a la caja a pagar). Del mismo modo, los intervalos de tiempo para completar una revisión de mantenimiento para un motor de automóvil o de avión poseen una distribución de frecuencia sesgada. La población asociada con estas variables aleatorias posee con frecuencia funciones de densidad que son modeladas de manera adecuada por una función de densidad gamma.

DEFINICIÓN **4.9**

Se dice que una variable aleatoria Y tiene una *distribución gamma con parámetros* $\alpha > 0$ y $\beta > 0$ si y sólo si la función de densidad de Y es

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & 0 \le y < \infty, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}\, dy.$$

La cantidad $\Gamma(\alpha)$ se conoce como *función gamma.* La integración directa verificará que $\Gamma(1) = 1$. La integración por partes verifica que $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ para cualquier $\alpha > 1$ y que $\Gamma(n) = (n - 1)!$, siempre que n sea un entero.

En la Figura 4.16 se dan gráficas de funciones de densidad gamma para $\alpha = 1$, 2 y 4 y $\beta = 1$. Observe en la Figura 4.16 que la forma de la densidad gamma difiere para los diferentes valores de α. Por esta razón, α recibe a veces el nombre de *parámetro de forma* asociado con una distribución gamma. El parámetro β generalmente se llama *parámetro de escala* porque multiplicar una variable aleatoria con distribución gamma por una constante positiva (y por tanto cambiando la escala en la que se hace la medición) produce una variable aleatoria

FIGURA **4.15** Función de densidad de probabilidad sesgada

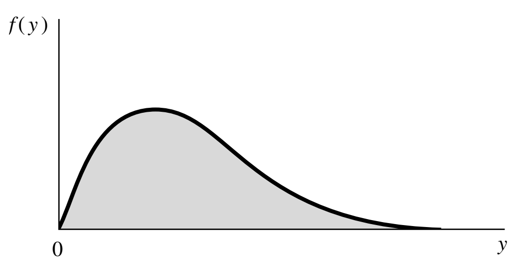
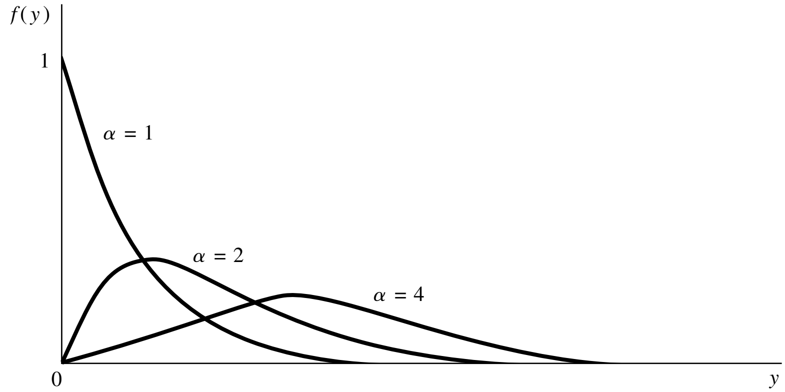


FIGURA **4.16** Funciones de densidad gamma, $\beta = 1$



que también tiene una distribución gamma con el mismo valor de α (parámetro de forma) pero con un valor alterado de β.

En el caso especial cuando α es un entero, la función de distribución de una variable aleatoria con distribución gamma puede expresarse como una suma de ciertas probabilidades de Poisson. Encontrará esta representación en el Ejercicio 4.99. Si α no es un entero y $0 < c < d < \infty$, es imposible dar una expresión de forma cerrada para

$$\int_c^d \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)}\, dy.$$

Por esta razón, excepto cuando $\alpha = 1$ (una distribución exponencial), es imposible obtener áreas bajo la función de densidad gamma por integración directa. Valores tabulados para integrales como ésta se dan en *Tables of the Incomplete Gamma Function* (Pearson 1965). Por mucho, la forma más fácil de calcular probabilidades asociadas con variables aleatorias de distribución gamma es usar un software de estadística. Si Y es una variable aleatoria con distribución gamma y parámetros α y β, el comando `pgamma(y0,α,1/β)` de *R* (o *S*–Plus) genera $P(Y \leq y_0)$, mientras que `qgamma(q,α,1/β)` da el p–ésimo cuantil, el valor de ϕ_p tal que $P(Y \leq \phi_p) = p$. Además, una de las aplicaciones breves, *Gamma Probabilities and Quantiles*, accesible en www.thomsonedu.com/statistics/wackerly, se puede usar para determinar probabilidades y cuantiles asociados con variables aleatorias de distribución gamma. Otra aplicación breve en la página web de Thomson, *Comparison of Gamma Density Functions,* permitirá visualizar y comparar funciones de densidad gamma con diferentes valores para α y/o β. Estas aplicaciones breves se usarán para contestar algunos de los ejercicios del final de esta sección.

Como se indica en el siguiente teorema, la media y la varianza de variables aleatorias de distribución gamma son fáciles de calcular.

TEOREMA **4.8**

Si Y tiene una distribución gamma con parámetros α y β, entonces

$$\mu = E(Y) = \alpha\beta \quad \text{y} \quad \sigma^2 = V(Y) = \alpha\beta^2.$$

Demostración

$$E(Y) = \int_{-\infty}^{\infty} yf(y)\,dy = \int_{0}^{\infty} y\left(\frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\right)dy.$$

Por definición, la función de densidad gamma es tal que

$$\int_{0}^{\infty} \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\,dy = 1.$$

Por tanto,

$$\int_{0}^{\infty} y^{\alpha-1}e^{-y\ \beta}\,dy = \beta^{\alpha}\Gamma(\alpha),$$

y

$$E(Y) = \int_{0}^{\infty} \frac{y^{\alpha}e^{-y/\beta}\,dy}{\beta^{\alpha}\Gamma(\alpha)} = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_{0}^{\infty} y^{\alpha}e^{-y/\beta}\,dy$$

$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}[\beta^{\alpha+1}\Gamma(\alpha+1)] = \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha\beta.$$

Del Ejercicio 4.24, $V(Y) = E[Y^2] - [E(Y)]^2$. Además,

$$E(Y^2) = \int_{0}^{\infty} y^2\left(\frac{y^{\alpha-1}e^{-y\ \beta}}{\beta^{\alpha}\Gamma(\alpha)}\right)dy = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_{0}^{\infty} y^{\alpha+1}e^{-y\ \beta}\,dy$$

$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}[\beta^{\alpha+2}\Gamma(\alpha+2)] = \frac{\beta^2(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha(\alpha+1)\beta^2.$$

Entonces $V(Y) = E[Y^2] - [E(Y)]^2$, donde, desde la primera parte de la derivación, $E(Y) = \alpha\beta$. Sustituyendo $E[Y^2]$ y $E(Y)$ en la fórmula para $V(Y)$, obtenemos

$$V(Y) = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

Dos casos especiales de variables aleatorias con distribución gamma ameritan consideración particular.

DEFINICIÓN **4.10**

Sea ν un entero positivo. Se dice que una variable aleatoria Y tiene *distribución ji cuadrada con* ν *grados de libertad* si y sólo si Y es una variable aleatoria con distribución gamma y parámetros $\alpha = \nu/2$ y $\beta = 2$.

Una variable aleatoria con distribución ji cuadrada se denomina *variable aleatoria* (χ^2) *ji cuadrada*. Estas variables aleatorias se presentan con frecuencia en teoría estadística. La motivación que hay detrás de llamar al parámetro ν como grados de libertad de la distribución χ^2 se apoya en una de las principales formas de generar una variable aleatoria con esta distribución y se da en el Teorema 6.4. La media y la varianza de una variable aleatoria χ^2 provienen directamente del Teorema 4.8.

TEOREMA **4.9**

Si Y es una variable aleatoria ji cuadrada con ν grados de libertad, entonces

$$\mu = E(Y) = \nu \quad \text{y} \quad \sigma^2 = V(Y) = 2\nu.$$

Demostración Aplique el Teorema 4.8 con $\alpha = \nu/2$ y $\beta = 2$.

En casi todos los textos de estadística se pueden ver tablas que dan probabilidades asociadas con distribuciones χ^2. La Tabla 6, Apéndice 3, da puntos porcentuales asociados con distribuciones χ^2 para numerosas opciones de ν. No se dispone fácilmente de tablas de la distribución gamma general, pero demostraremos en el Ejercicio 6.46 que si Y tiene una distribución gamma con $\alpha = n/2$ para algún entero n, entonces $2Y/\beta$ tiene una distribución χ^2 con n grados de libertad. De ahí que, por ejemplo, si Y tiene una distribución gamma con $\alpha = 1.5 = 3/2$ y $\beta = 4$, entonces $2Y/\beta = 2Y/4 = Y/2$ tiene una distribución χ^2 con 3 grados de libertad. Entonces, $P(Y < 3.5) = P([Y/2] < 1.75)$ se puede hallar usando tablas de la distribución χ^2 de las que se puede disponer fácilmente.

La función de densidad gamma en la que $\alpha = 1$, se llama *función de densidad exponencial.*

DEFINICIÓN **4.11**

Se dice que una variable aleatoria Y tiene una *distribución exponencial con parámetro* $\beta > 0$ si y sólo si la función de densidad de Y es

$$f(y) = \begin{cases} \dfrac{1}{\beta} e^{-y/\beta}, & 0 \le y < \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La función de densidad exponencial a menudo es de ayuda para modelar la vida útil de componentes electrónicos. Suponga que el tiempo que ya ha operado un componente no afecta su probabilidad de operar durante al menos b unidades de tiempo adicionales. Esto es, la probabilidad de que el componente opere durante más de $a + b$ unidades de tiempo, dado que ya ha operado durante al menos a unidades de tiempo, es la misma que la probabilidad de que un componente nuevo opere al menos b unidades de tiempo si el componente nuevo se pone en servicio en el tiempo 0. Un fusible es un ejemplo de un componente para el cual a veces esta suposición es razonable. Veremos en el siguiente ejemplo que la distribución exponencial proporciona un modelo para la distribución de la vida útil de ese componente.

TEOREMA **4.10**

Si Y es una variable aleatoria exponencial con parámetro β, entonces

$$\mu = E(Y) = \beta \quad \text{y} \quad \sigma^2 = V(Y) = \beta^2.$$

Demostración La demostración se sigue directamente del Teorema 4.8 con $\alpha = 1$.

EJEMPLO **4.10** Suponga que Y tiene una función de densidad de probabilidad exponencial. Demuestre que, si $a > 0$ y $b > 0$,

$$P(Y > a + b | Y > a) = P(Y > b).$$

Solución De la definición de probabilidad condicional, tenemos que

$$P(Y > a + b|Y > a) = \frac{P(Y > a + b)}{P(Y > a)}$$

porque la intersección de los eventos $(Y > a + b)$ y $(Y > a)$ es el evento $(Y > a + b)$. Ahora

$$P(Y > a + b) = \int_{a+b}^{\infty} \frac{1}{\beta} e^{-y/\beta}\, dy = -e^{-y/\beta}\Big]_{a+b}^{\infty} = e^{-(a+b)/\beta}.$$

De manera similar,

$$P(Y > a) = \int_{a}^{\infty} \frac{1}{\beta} e^{-y/\beta}\, dy = e^{-a/\beta},$$

y

$$P(Y > a + b|Y > a) = \frac{e^{-(a+b)/\beta}}{e^{-a/\beta}} = e^{-b/\beta} = P(Y > b).$$

Esta propiedad de la distribución exponencial en ocasiones recibe el nombre de *propiedad sin memoria* de la distribución. ■

Como recordará del Capítulo 3, la distribución geométrica, que es una distribución discreta, también tenía propiedad *sin memoria.* Una relación interesante entre las distribuciones exponencial y geométrica se da en el Ejercicio 4.95.

Ejercicios

4.81 **a** Si $\alpha > 0$, $\Gamma(\alpha)$ está definida por $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\, dy$, demuestre que $\Gamma(1) = 1$.

*__b__ Si $\alpha > 1$, integre por partes para demostrar que $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

4.82 Use los resultados obtenidos en el Ejercicio 4.81 para demostrar que si n es un entero positivo, entonces $\Gamma(n) = (n - 1)!$. ¿Cuáles son los valores numéricos de $\Gamma(2)$, $\Gamma(4)$ y $\Gamma(7)$?

4.83 **Ejercicio Applet** Use la aplicación breve *Comparison of Gamma Density Functions* para obtener los resultados dados en la Figura 4.16.

4.84 **Ejercicio Applet** Consulte el Ejercicio 4.83. Use la aplicación breve *Comparison of Gamma Density Functions* para comparar funciones con densidad gamma con $(\alpha = 4,\ \beta = 1)$, $(\alpha = 40,\ \beta = 1)$ y $(\alpha = 80,\ \beta = 1)$.

a ¿Qué observa usted acerca de las formas de estas tres funciones de densidad? ¿Cuáles son menos sesgadas y más simétricas?

b ¿Qué diferencias observa acerca de la ubicación de los centros de estas funciones de densidad?

c Dé una explicación de lo que observó en el inciso b.

4.85 **Ejercicio Applet** Use la aplicación breve *Comparison of Gamma Density Functions* para comparar funciones de densidad gamma con ($\alpha = 1,\ \beta = 1$), ($\alpha = 1,\ \beta = 2$) y ($\alpha = 1,\ \beta = 4$).

a ¿Qué otro nombre se da a las funciones de densidad que observó?

b ¿Estas densidades tienen la misma forma general?

c El parámetro β es un parámetro "de escala". ¿Qué observa usted acerca de la "dispersión" de estas tres funciones de densidad?

4.86 **Ejercicio Applet** Cuando examinamos la distribución χ^2 en esta sección, presentamos (con justificación en el Capítulo 6) el hecho de que si Y tiene distribución gamma con $\alpha = n/2$ para algún entero n, entonces $2Y/\beta$ tiene una distribución χ^2. En particular, se dijo que cuando $\alpha = 1.5$ y $\beta = 4$, $W = Y/2$ tiene una distribución χ^2 con tres grados de libertad.

a Use la aplicación *Gamma Probabilities and Quantiles* para hallar $P(Y < 3.5)$.

b Use la aplicación *Gamma Probabilities and Quantiles* para hallar $P(W < 1.75)$. [*Sugerencia:* Recuerde que la distribución χ^2 con ν grados de libertad es simplemente una distribución gamma con $\alpha = \nu/2$ y $\beta = 2$.]

c Compare sus respuestas a los incisos a y b.

4.87 **Ejercicio Applet** Hagamos que Y y W tengan las distribuciones dadas en el Ejercicio 4.86.

a Use la aplicación *Gamma Probabilities and Quantiles* para hallar el cuantil .05 de la distribución de Y.

b Use la aplicación *Gamma Probabilities and Quantiles* para hallar el cuantil .05 de la distribución χ^2 con tres grados de libertad.

c ¿Cuál es la relación entre el cuantil .05 de la distribución gamma ($\alpha = 1.5, \beta = 4$) y el cuantil .05 de la distribución χ^2 con tres grados de libertad? Explique esta relación.

4.88 La magnitud de temblores registrados en una región de América del Norte puede modelarse como si tuviera una distribución exponencial con media 2.4, según se mide en la escala de Richter. Encuentre la probabilidad de que un temblor que ocurra en esta región

a sea mayor que 3.0 en la escala de Richter.

b caiga entre 2.0 y 3.0 en la escala de Richter.

4.89 Si Y tiene una distribución exponencial y $P(Y > 2) = .0821$, ¿cuál es

a $\beta = E(Y)$?

b $P(Y \leq 1.7)$?

4.90 Consulte el Ejercicio 4.88. De los siguientes diez temblores que afecten esta región, ¿cuál es la probabilidad de que al menos uno de ellos sea mayor que 5.0 en la escala de Richter?

4.91 El operador de una estación de bombeo ha observado que la demanda de agua durante las primeras horas de la tarde tiene una distribución aproximadamente exponencial con media de 100 pes (pcs cúbicos por segundo).

a Encuentre la probabilidad de que la demanda sea mayor que 200 pcs durante las primeras horas de la tarde en un día seleccionado al azar.

b ¿Qué capacidad de bombeo de agua debe mantener la estación durante las primeras horas de la tarde para que la probabilidad de que la demanda sea mayor que la capacidad en un día seleccionado al azar sea de sólo .01?

4.92 El tiempo Y necesario para completar una operación clave en la construcción de casas tiene una distribución exponencial con media de 10 horas. La fórmula $C = 100 + 40Y + 3Y^2$ relaciona el costo C de completar esta operación con el cuadrado del tiempo para completarla. Encuentre la media y la varianza de C.

4.93 Una evidencia histórica indica que los tiempos entre accidentes mortales en vuelos nacionales de horario programado en aviones de pasajeros en Estados Unidos tienen una distribución aproximadamente exponencial. Suponga que el tiempo medio entre accidentes es de 44 días.

a Si uno de los accidentes ocurrió el 1 de julio de un año seleccionado al azar en el periodo de estudio, ¿cuál es la probabilidad de que ocurra otro accidente ese mismo mes?

b ¿Cuál es la varianza de los tiempos entre accidentes?

4.94 Concentraciones de monóxido de carbono de una hora en muestras de aire de una gran ciudad tienen una distribución aproximadamente exponencial, con media de 3.6 ppm (partes por millón).

a Encuentre la probabilidad de que la concentración de monóxido de carbono exceda de 9 ppm durante un periodo de una hora seleccionado al azar.

b Una estrategia de control de tránsito redujo la media a 2.5 ppm. Ahora encuentre la probabilidad de que la concentración exceda de 9 ppm.

4.95 Sea Y una variable aleatoria distribuida exponencialmente con media β. Defina una variable aleatoria X en la siguiente forma: $X = k$ si $k - 1 \leq Y < k$ para $k = 1, 2, \ldots$

a Encuentre $P(X = k)$ para cada $k = 1, 2, \ldots$

b Demuestre que su respuesta al inciso a se puede escribir como

$$P(X = k) = \left(e^{-1/\beta}\right)^{k-1}\left(1 - e^{-1/\beta}\right), \qquad k = 1,\ 2, \ldots$$

y que X tiene una distribución geométrica con $p = (1 - e^{-1/\beta})$.

4.96 Suponga que una variable aleatoria Y tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} ky^3e^{-y/2}, & y > 0. \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de k que haga de $f(y)$ una función de densidad.

b ¿Tiene Y una distribución χ^2? Si es así, ¿de cuántos grados de libertad?

c ¿Cuáles son la media y la desviación estándar de Y?

d **Ejercicio Applet** ¿Cuál es la probabilidad de que Y se encuentre a no más de 2 desviaciones estándar de su media?

4.97 Una planta de manufactura utiliza un producto específico a granel. La cantidad de producto empleada en un día puede ser modelada por una distribución exponencial con $\beta = 4$ (medida en toneladas). Encuentre la probabilidad de que la planta utilice más de 4 toneladas en un día determinado.

4.98 Considere la planta del Ejercicio 4.97. ¿Cuánto producto a granel debe tener en existencia para que la probabilidad de que se agote el producto en la planta sea de sólo .05?

4.99 Si $\lambda > 0$ y α es un entero positivo, la relación entre integrales gamma incompletas y sumas de probabilidades de Poisson está dada por

$$\frac{1}{\Gamma(\alpha)} \int_{\lambda}^{\infty} y^{\alpha-1}e^{-y}\, dy = \sum_{x=0}^{\alpha-1} \frac{\lambda^x e^{-\lambda}}{x!}.$$

a Si Y tiene una distribución gamma con $\alpha = 2$ y $\beta = 1$, encuentre $P(Y > 1)$ usando la igualdad anterior y la Tabla 3 del Apéndice 3.

b **Ejercicio Applet** Si Y tiene una distribución gamma con $\alpha = 2$ y $\beta = 1$, encuentre $P(Y > 1)$ con el uso de la aplicación breve *Gamma Probabilities.*

***4.100** Sea Y una variable aleatoria con distribución gamma donde α es un entero positivo y $\beta = 1$. El resultado dado en el Ejercicio 4.99 implica que si $y > 0$,

$$\sum_{x=0}^{\alpha-1} \frac{y^x e^{-y}}{x!} = P(Y > y).$$

Suponga que X_1 tiene distribución de Poisson con media λ_1 y X_2 tiene distribución de Poisson con media λ_2, donde $\lambda_2 > \lambda_1$.

a Demuestre que $P(X_1 = 0) > P(X_2 = 0)$.

b Sea k cualquier entero positivo fijo. Demuestre que $P(X_1 \leq k) = P(Y > \lambda_1)$ y $P(X_2 \leq k) = P(Y > \lambda_2)$, donde Y es una distribución gamma con $\alpha = k + 1$ y $\beta = 1$.

c Sea k cualquier entero fijo positivo. Use el resultado que obtuvo en el inciso b y el hecho de que $\lambda_2 > \lambda_1$ para demostrar que $P(X_1 \leq k) > P(X_2 \leq k)$.

d Como el resultado del inciso c es válido para cualquier $k = 1, 2, 3, \ldots$ y el inciso a también es válido, hemos establecido que $P(X_1 \leq k) > P(X_2 \leq k)$ para toda $k = 0, 1, 2\ldots$ Interprete este resultado.

4.101 **Ejercicio Applet** Consulte el Ejercicio 4.88. Suponga que la magnitud de los terremotos que afectan la región tiene una distribución gamma con $\alpha = .8$ y $\beta = 2.4$.

a ¿Cuál es la magnitud media de los terremotos que afectan la región?

b ¿Cuál es la probabilidad de que la magnitud de un terremoto que afecte la región exceda de 3.0 en la escala Richter?

c Compare sus respuestas con el Ejercicio 4.88(a). ¿Cuál probabilidad es mayor? Explique.

d ¿Cuál es la probabilidad de que un terremoto que afecte las regiones caiga entre 2.0 y 3.0 en la escala Richter?

4.102 **Ejercicio Applet** Consulte el Ejercicio 4.97. Suponga que la cantidad de producto usado en un día tiene una distribución gamma con $\alpha = 1.5$ y $\beta = 3$.

a Encuentre la probabilidad de que la planta use más de 4 toneladas en un día determinado.

b ¿Cuánto del producto a granel debe haber en existencia para que la probabilidad de que la planta agote el producto sea de sólo .05?

4.103 Materiales explosivos que se usan en operaciones de minería producen cráteres casi circulares cuando se hacen detonar. Los radios de estos cráteres están distribuidos exponencialmente con media de 10 pies. Encuentre la media y la varianza de las áreas producidas por estos materiales explosivos.

4.104 La vida útil (en horas) Y de un componente electrónico es una variable aleatoria con función de densidad dada por

$$f(y) = \begin{cases} \frac{1}{100} e^{-y/100}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Tres componentes operan de manera independiente en una pieza de equipo. El equipo falla si fallan al menos dos de los componentes. Encuentre la probabilidad de que el equipo opere durante al menos 200 horas sin fallar.

4.105 La cantidad total de lluvia de cuatro semanas en verano en una parte del medio oeste de Estados Unidos tiene aproximadamente una distribución gamma con $\alpha = 1.6$ y $\beta = 2.0$.

a Encuentre la media y varianza del total de lluvia de cuatro semanas.

b **Ejercicio Applet** ¿Cuál es la probabilidad de que la cantidad total de lluvia de cuatro semanas sea mayor que 4 pulgadas?

4.106 Los tiempos de respuesta en una terminal de computadora en línea tienen aproximadamente una distribución gamma con media de cuatro segundos y varianza de ocho segundos^2.

a Escriba una función de densidad de probabilidad para los tiempos de respuesta.

b **Ejercicio Applet** ¿Cuál es la probabilidad de que el tiempo de respuesta en la terminal sea menor que cinco segundos?

4.107 Consulte el Ejercicio 4.106.

a Use el teorema de Tchebysheff para dar un intervalo que contenga al menos 75% de los tiempos de respuesta.

b **Ejercicio Applet** ¿Cuál es la probabilidad real de observar un tiempo de respuesta en el intervalo que obtuvo en el inciso a?

4.108 Los ingresos anuales de jefes de familia en una parte de una ciudad tienen aproximadamente una distribución gamma con $\alpha = 20$ y $\beta = 1000$.

a Encuentre la media y la varianza de estos ingresos.

b ¿Esperaría hallar muchos ingresos de más de \$30,000 en esta sección de la ciudad?

c **Ejercicio Applet** ¿Qué proporción de jefes de familia de esta sección de la ciudad tienen ingresos de más de \$30,000?

4.109 El tiempo improductivo por semana Y (en horas) de una máquina industrial tiene aproximadamente una distribución gamma con $\alpha = 3$ y $\beta = 2$. La pérdida L (en dólares) para la operación industrial como resultado de este tiempo improductivo está dada por $L = 30Y + 2Y^2$. Encuentre la varianza y el valor esperados de L.

4.110 Si Y tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 4y^2e^{-2y}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

obtenga $E(Y)$ y $V(Y)$ por inspección.

4.111 Suponga que Y tiene una distribución gamma con parámetros α y β.

a Si a es cualquier valor positivo o negativo tal que $\alpha + a > 0$, demuestre que

$$E(Y^a) = \frac{\beta^a \Gamma(\alpha + a)}{\Gamma(\alpha)}.$$

b ¿Por qué la respuesta en el inciso a requirió que $\alpha + a > 0$?

c Demuestre que, con $a = 1$, el resultado del inciso a da $E(Y) = \alpha\beta$.

d Use el resultado del inciso a para dar una expresión para $E(\sqrt{Y})$. ¿Qué es necesario suponer acerca de α?

e Utilice el resultado del inciso a para obtener una expresión para $E(1/Y)$, E($/\sqrt{y}$) y E($1/Y^2$). ¿Qué es necesario suponer sobre α en cada caso?

4.112 Suponga que Y tiene una distribución χ^2 con ν grados de libertad. Use los resultados del Ejercicio 4.111 en sus respuestas a lo siguiente. Estos resultados serán útiles cuando estudiemos las distribuciones t y F en el Capítulo 7.

a Proporcione una expresión para $E(Y^a)$ si $\nu > -2a$.

b ¿Por qué la respuesta en el inciso a requirió que $\nu > -2a$?

c Use el resultado del inciso a para dar una expresión para $E(\sqrt{Y})$. ¿Qué es necesario suponer acerca de ν?

d Use el resultado del inciso a para dar una expresión para $E(1/Y)$, $E(1/\sqrt{Y})$ y $E(1/Y^2)$. ¿Qué es necesario suponer acerca de ν en cada caso?

4.7 La distribución de probabilidad beta

La función de densidad beta es una función de densidad de dos parámetros definida sobre el intervalo cerrado $0 \le y \le 1$. Frecuentemente se usa como modelo para proporciones, por ejemplo como la proporción de impurezas en un producto químico o la proporción de tiempo que una máquina está en reparación.

DEFINICIÓN **4.12**

Se dice que una variable aleatoria Y tiene una *distribución de probabilidad beta con parámetros* $\alpha > 0$ *y* $\beta > 0$ si y sólo si la función de densidad de Y es

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Las gráficas de funciones de densidad beta toman formas muy diferentes para diversos valores de los dos parámetros α y β. Algunos de éstos se muestran en la Figura 4.17. Ciertos ejercicios del final de esta sección piden al lector usar la applet *Comparison of Beta Density Functions* accesibles en www.thomsonedu.com/statistics/wackerly para explorar y comparar las formas de más densidades beta.

Observe que definir y sobre el intervalo $0 \le y \le 1$ no restringe el uso de la distribución beta. Si $c \le y \le d$, entonces $y^* = (y - c)/(d - c)$ define una nueva variable tal que $0 \le y^* \le 1$. Entonces, la función de densidad beta se puede aplicar a una variable aleatoria definida en el intervalo $c \le y \le d$ por traducción y un cambio de escala.

La función de distribución acumulativa para la variable aleatoria beta comúnmente se denomina *función beta incompleta* y está denotada por

$$F(y) = \int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}\,dt = I_y(\alpha,\beta)$$

Una tabulación de $I_y(\alpha, \beta)$ se da en la obra *Tables of the Incomplete Beta Function* (Pearson, 1968). Cuando α y β son enteros positivos $I_y(\alpha, \beta)$ está relacionada con la función de proba-

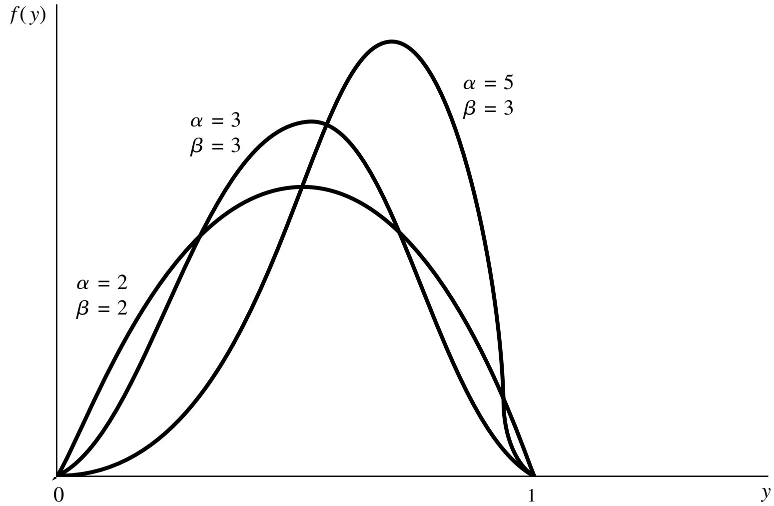


FIGURA **4.17** Funciones de densidad beta

bilidad binomial. Es posible usar integración por partes para demostrar que $0 < y < 1$, y α y β ambos enteros,

$$F(y) = \int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}\,dt = \sum_{i=\alpha}^{n} \binom{n}{i} y^i (1-y)^{n-i},$$

donde $n = \alpha + \beta - 1$. Observe que la suma del lado derecho de esta expresión es precisamente la suma de probabilidades asociadas con una variable aleatoria binomial $n = \alpha + \beta - 1$ y $p = y$. La función de distribución acumulativa binomial se presenta en la Tabla 1, Apéndice 3, para $n = 5, 10, 15, 20$ y 25 y $p = .01, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95$ y .99. El modo más eficiente de obtener probabilidades binomiales es usar un software de estadística como el *R* o *S*–Plus (vea el Capítulo 3). Una forma incluso más fácil para hallar probabilidades y cuantiles asociados con variables aleatorias de distribución beta es usar directamente software apropiado. La página web de Thomson contiene una aplicación breve, *Beta Probabilities,* que proporciona probabilidades de "cola superior" [es decir, $P(Y > y_0)$] y cuantiles asociados con variables aleatorias con distribución beta. Además, si Y es una variable aleatoria con distribución beta y parámetros α y β, el comando `pbeta(`y_0, α, $1/\beta$`)` de *R* (o *S*–Plus) genera $P(Y \le y_0)$, mientras que `qbeta(p,` α, $1/\beta$`)` da el p–ésimo cuantil, el valor de ϕ_p de manera que $P(Y \le \phi_p) = p$.

TEOREMA **4.11**

Si Y es una variable aleatoria con distribución beta $\alpha > 0$ y $\beta > 0$, entonces

$$\mu = E(Y) \quad \frac{\alpha}{\alpha+\beta} \quad \text{y} \quad \sigma^2 = V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Demostración Por definición,

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f(y)\, dy \\
&= \int_{0}^{1} y \left[\frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)}\right] dy \\
&= \frac{1 \quad \beta}{B(\alpha, \beta)} \int_{0}^{1} y^{\alpha}(1-y)^{\beta-1}\, dy \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \qquad \text{(porque } \alpha > 0 \text{ implica que } \alpha + 1 > 0) \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{(\alpha+\beta)}.
\end{aligned}
$$

La obtención de la varianza se deja al lector (vea Ejercicio 4.130).

Veremos en el siguiente ejemplo que la función de densidad beta se puede integrar directamente cuando α y β son enteros.

EJEMPLO **4.11** Una distribuidora mayorista de gasolina tiene tanques de almacenamiento a granel que contienen suministros fijos y se llenan cada lunes. De interés para la mayorista es la proporción de este suministro que se vende durante la semana. Durante varias semanas de observación, la distribuidora encontró que esta proporción podría ser modelada por una distribución beta con $\alpha = 4$ y $\beta = 2$. Encuentre la probabilidad de que la mayorista venda al menos 90% de su existencia en una semana determinada.

Solución Si Y denota la proporción vendida durante la semana, entonces

$$
f(y) = \begin{cases} \dfrac{\Gamma(4+2)}{\Gamma(4)\Gamma(2)} y^3(1-y), & 0 \le y \le 1, \\[2ex] 0, & \text{en otro lugar,} \end{cases}
$$

y

$$
\begin{aligned}
P(Y > .9) &= \int_{9}^{\infty} f(y)\, dy = \int_{9}^{1} 20(y^3 - y^4)\, dy \\
&= 20\left\{ \frac{y^4}{4}\Big]_{.9}^{1} - \frac{y^5}{5}\Big]_{.9}^{1} \right\} = 20(.004) = .08.
\end{aligned}
$$

No es muy probable que 90% de la existencia se venda en una semana determinada. ■

Ejercicios

4.113 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para obtener los resultados dados en la Figura 4.17.

4.114 **Ejercicio Applet** Consulte el Ejercicio 4.113. Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = 1,\ \beta = 1)$, $(\alpha = 1,\ \beta = 2)$ y $(\alpha = 2,\ \beta = 1)$.

a ¿Cómo hemos llamado previamente a la distribución beta con $(\alpha = 1, \beta = 1)$?

b ¿Cuál de estas densidades beta es simétrica?

c ¿Cuál de estas densidades beta está sesgada a la derecha?

d ¿Cuál de estas densidades beta está sesgada a la izquierda?

***e** En el Capítulo 6 veremos que si Y tiene distribución beta con parámetros α y β, entonces $Y^* = 1 - Y$ tiene una distribución beta con parámetros $\alpha^* = \beta$ y $\beta^* = \alpha$. ¿Esto explica las diferencias en las gráficas de las densidades beta?

4.115 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = 2,\ \beta = 2)$, $(\alpha = 3,\ \beta = 3)$ y $(\alpha = 9,\ \beta = 9)$.

a ¿Cuáles son las medias asociadas con las variables aleatorias con cada una de estas distribuciones beta?

b ¿Qué es semejante acerca de estas densidades?

c ¿Cómo difieren estas densidades? En particular, ¿qué observa usted acerca de la "dispersión" de estas tres funciones de densidad?

d Calcule las desviaciones estándar asociadas con las variables aleatorias con cada una de estas densidades beta. ¿Los valores de estas desviaciones estándar explican lo que se observó en el inciso c? Explique.

e Grafique algunas densidades beta adicionales con $\alpha = \beta$. ¿Qué se puede conjeturar acerca de la forma de las densidades beta con $\alpha = \beta$?

4.116 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = 1.5,\ \beta = 7)$, $(\alpha = 2.5,\ \beta = 7)$ y $(\alpha = 3.5,\ \beta = 7)$.

a ¿Son simétricas estas densidades? ¿Sesgadas a la izquierda? ¿Sesgadas a la derecha?

b ¿Qué observa usted a medida que el valor de α se acerca a 7?

c Grafique algunas densidades beta adicionales con $\alpha > 1$, $\beta > 1$, y $\alpha < \beta$. ¿Qué puede conjeturar acerca de la forma de las densidades beta cuando $\alpha > 1$, $\beta > 1$ y $\alpha < \beta$?

4.117 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = 9, \beta = 7)$, $(\alpha = 10, \beta = 7)$ y $(\alpha = 12, \beta = 7)$.

a ¿Son simétricas estas densidades? ¿Sesgadas a la izquierda? ¿Sesgadas a la derecha?

b ¿Qué observa usted a medida que el valor de α se acerca a 12?

c Grafique algunas densidades beta adicionales con $\alpha > 1$, $\beta > 1$ y $\alpha > \beta$. ¿Qué se puede conjeturar acerca de la forma de las densidades beta con $\alpha > \beta$ y $\alpha > 1$ y $\beta > 1$?

4.118 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = .3, \beta = 4)$, $(\alpha = .3, \beta = 7)$ y $(\alpha = .3, \beta = 12)$.

a ¿Son simétricas estas densidades? ¿Sesgadas a la izquierda? ¿Sesgadas a la derecha?

b ¿Qué observa usted a medida que el valor de β se acerca a 12?

c ¿Cuál de estas distribuciones beta da la más alta probabilidad de observar un valor mayor que .2?

d Grafique algunas densidades beta adicionales con $\alpha < 1$ y $\beta > 1$. ¿Qué se puede conjeturar acerca de la forma de las densidades beta con $\alpha < 1$ y $\beta > 1$?

4.119 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = 4,\ \beta = .3)$, $(\alpha = 7,\ \beta = .3)$, y $(\alpha = 12,\ \beta = .3)$.

a ¿Son simétricas estas densidades? ¿Sesgadas a la izquierda? ¿Sesgadas a la derecha?

b ¿Qué observa usted cuando el valor de α se acerca a 12?

c ¿Cuál de estas distribuciones beta da la más alta probabilidad de observar un valor menor que .8?

d Grafique algunas densidades beta adicionales con $\alpha > 1$ y $\beta < 1$. ¿Qué se puede conjeturar acerca de la forma de las densidades beta con $\alpha > 1$ y $\beta < 1$?

***4.120** En el Capítulo 6 veremos que si Y tiene distribución beta con parámetros α y β, entonces $Y^* = 1 - Y$ tiene una distribución beta con parámetros $\alpha^* = \beta$ y $\beta^* = \alpha$. ¿Explica esto las diferencias y similitudes en las gráficas de las densidades beta en los Ejercicios 4.118 y 4.119?

4.121 **Ejercicio Applet** Use la aplicación breve *Comparison of Beta Density Functions* para comparar funciones de densidad beta con $(\alpha = .5,\ \beta = .7)$, $(\alpha = .7,\ \beta = .7)$ y $(\alpha = .9,\ \beta = .7)$.

a ¿Cuál es la forma general de estas densidades?

b ¿Qué observa cuando el valor de α aumenta?

4.122 **Ejercicio Applet** Las densidades beta con $\alpha < 1$ y $\beta < 1$ son difíciles de exhibir debido a problemas de escala o resolución.

a Use la aplicación breve *Beta Probabilities and Quantiles* para calcular $P(Y > .1)$ si Y tiene una distribución beta con $(\alpha = .1, \beta = 2)$.

b Use la aplicación breve *Beta Probabilities and Quantiles* para calcular $P(Y < .1)$ si Y tiene una distribución beta con $(\alpha = .1, \beta = 2)$.

c Con base en su respuesta el inciso b, ¿a cuáles valores de Y se les asignan altas probabilidades si Y tiene una distribución beta con $(\alpha = .1, \beta = 2)$?

d Use la aplicación breve *Beta Probabilities and Quantiles* para calcular $P(Y < .1)$ si Y tiene una distribución beta con $(\alpha = .1, \beta = .2)$.

e Use la aplicación breve *Beta Probabilities and Quantiles* para calcular $P(Y > .9)$ si Y tiene una distribución beta con $(\alpha = .1, \beta = .2)$.

f Use la aplicación breve *Beta Probabilities and Quantiles* para calcular $P(.1 < Y < .9)$ si Y tiene una distribución beta con $(\alpha = 1, \beta = .2)$.

g Con base en sus respuestas a los incisos d, e y f, ¿a cuáles valores de Y se les asignan altas probabilidades si Y tiene una distribución beta con $(\alpha = .1, \beta = .2)$?

4.123 La humedad relativa Y, cuando se mide en una localidad, tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} ky^3(1-y)^2, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de k que haga de $f(y)$ una función de densidad.

b **Ejercicio Applet** Use la aplicación breve *Beta Probabilities and Quantiles* para hallar un valor de humedad que se exceda sólo 5% del tiempo.

4.124 El porcentaje de impurezas por lote en un producto químico es una variable aleatoria Y con función de densidad

$$f(y) = \begin{cases} 12y^2(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Un lote con más de 40% de impurezas no se puede vender.

a Integre la densidad directamente para determinar la probabilidad de que un lote seleccionado al azar no se pueda vender por su exceso de impurezas.

b **Ejercicio Applet** Use la aplicación breve *Beta Probabilities and Quantiles* para hallar la respuesta al inciso a.

4.125 Consulte el Ejercicio 4.124. Encuentre la media y la varianza del porcentaje de impurezas en un lote seleccionado al azar del producto químico.

4.126 Suponga que una variable aleatoria Y tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 6y(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $F(y)$.

b Grafique $F(y)$ y $f(y)$.

c Encuentre $P(.5 \le Y \le .8)$.

4.127 Verifique que si Y tiene una distribución beta con $\alpha = \beta = 1$, entonces Y tiene una distribución uniforme en el intervalo (0, 1). Esto es, la distribución uniforme en el intervalo (0, 1) es un caso especial de una distribución beta.

4.128 El costo Y de reparaciones semanales para una máquina tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 3(1-y)^2, & 0 < y < 1, \\ 0. & \text{en cualquier otro punto,} \end{cases}$$

con mediciones en cientos de dólares. ¿Cuánto dinero debe presupuestarse a la semana para costos de reparación, de modo que el costo real rebase la cantidad presupuestada sólo 10% del tiempo?

4.129 Durante un turno de ocho horas la proporción de tiempo Y que una máquina troqueladora de láminas metálicas está sin operar por mantenimiento o reparaciones tiene una distribución beta con $\alpha = 1$ y $\beta = 2$. Esto es,

$$f(y) = \begin{cases} 2(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El costo (en cientos de dólares) de este tiempo improductivo, debido a producción perdida y costo de mantenimiento y reparación, está dado por $C = 10 + 20Y + 4Y^2$. Encuentre la media y la varianza de C.

4.130 Demuestre que la varianza de una variable aleatoria con distribución beta y parámetros α y β es

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

4.131 Errores en la medición del tiempo de llegada de un frente de onda, desde una fuente acústica, en ocasiones tienen una distribución beta aproximada. Suponga que estos errores, medidos en microsegundos, tienen aproximadamente una distribución beta con $\alpha = 1$ y $\beta = 2$.

a ¿Cuál es la probabilidad de que el error de medición en un ejemplo seleccionado al azar sea menor que .5 μs?

b Obtenga la media y la desviación estándar de los errores de medición.

4.132 La mezcla apropiada de polvos finos y gruesos antes de la sinterización (aglutinación) del cobre es esencial para lograr uniformidad en el producto acabado. Un modo de verificar la homogeneidad de la mezcla es seleccionar muchas muestras pequeñas de los polvos mezclados y medir la proporción del peso total aportado por los polvos finos en cada una. Estas mediciones deben ser relativamente estables si se ha obtenido una mezcla homogénea.

a Suponga que la proporción del peso total aportada por los polvos finos tiene una distribución beta con $\alpha = \beta = 3$. Encuentre la media y la varianza de la proporción de peso aportada por los polvos finos.

b Repita el inciso a si $\alpha = \beta = 2$.

c Repita el inciso a si $\alpha = \beta = 1$.

d ¿Cuál de los casos —incisos a, b o c— dan la mezcla más homogénea?

4.133 La proporción de tiempo por día que todas las cajas en un supermercado están ocupadas es una variable aleatoria Y con una función de densidad dada por

$$f(y) = \begin{cases} cy^2(1-y)^4, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de c que haga de $f(y)$ una función de densidad de probabilidad.

b Encuentre $E(Y)$. (Use lo que haya aprendido acerca de la distribución tipo beta. Compare sus respuestas con las obtenidas en el Ejercicio 4.28.)

c Calcule la desviación estándar de Y.

d **Ejercicio Applet** Utilice la aplicación breve *Beta Probabilities and Quantiles* para hallar $P(Y > \mu + 2\sigma)$.

4.134 En el texto de esta sección observamos la relación entre la función de distribución de una variable aleatoria con distribución beta y las sumas de probabilidades binomiales. Específicamente, si Y tiene una distribución beta con valores enteros positivos para α y β y $0 < y < 1$,

$$F(y) = \int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)}\, dt = \sum_{i=\alpha}^{n} \binom{n}{i} y^i (1-y)^{n-i},$$

donde $n = \alpha + \beta - 1$.

a Si Y tiene una distribución beta con $\alpha = 4$ y $\beta = 7$, use las tablas binomiales apropiadas para hallar $P(Y \le .7) = F(.7)$.

b Si Y tiene una distribución beta con $\alpha = 12$ y $\beta = 14$, use las tablas binomiales apropiadas para hallar $P(Y \le .6) = F(.6)$.

c **Ejercicio Applet** Use la aplicación breve *Beta Probabilities and Quantiles* para hallar las probabilidades en los incisos a y b.

***4.135** Suponga que Y_1 y Y_2 son variables aleatorias binomiales con parámetros (n, p_1) y (n, p_2), respectivamente, donde $p_1 < p_2$. (Observe que el parámetro n es el mismo para las dos variables.)

a Use la fórmula binomial para deducir que $P(Y_1 = 0) > P(Y_2 = 0)$.

b Use la relación entre la función de distribución beta y las sumas de probabilidades binomiales dadas en el Ejercicio 4.134 para deducir que, si k es un entero entre 1 y $n - 1$,

$$P(Y_1 \le k) = \sum_{i=0}^{k} \binom{n}{i} (p_1)^i (1-p_1)^{n-i} = \int_{p_1}^{1} \frac{t^k (1-t)^{n-k-1}}{B(k+1, n-k)}\, dt.$$

c Si k es un entero entre 1 y n – 1, el mismo argumento empleado en el inciso b dice que

$$P(Y_2 \le k) = \sum_{i=0}^{k} \binom{n}{i} (p_2)^i (1 - p_2)^{n-i} = \int_{p_2}^{1} \frac{t^k (1-t)^{n-k-1}}{B(k+1, n-k)}\, dt.$$

Demuestre que si k es cualquier entero entre 1 y n – 1, $P(Y_1 \le k) > P(Y_2 \le k)$. Interprete este resultado.

4.8 Algunos comentarios generales

Recuerde que las funciones de densidad son modelos teóricos para poblaciones de datos reales que se presentan en fenómenos aleatorios. ¿Cómo sabemos cuál modelo usar? ¿Cuánto importa si usamos la densidad errónea como modelo para la realidad?

Para contestar primero esta última pregunta, es muy poco probable que alguna vez seleccionemos una función de densidad que proporcione una representación perfecta de la naturaleza; pero la bondad de ajuste no es el criterio para evaluar lo adecuado de nuestro modelo. La finalidad de un modelo probabilístico es proveer el mecanismo para hacer inferencias acerca de una población con base en información contenida en una muestra. La probabilidad de la muestra observada (o una cantidad proporcional a ella) es útil para hacer una inferencia acerca de la población. Se deduce que una función de densidad que proporcione un mal ajuste a la distribución de frecuencia poblacional (pero no necesariamente) da como resultado enunciados incorrectos de probabilidad y lleva a inferencias erróneas acerca de la población. Un buen modelo es aquel que produzca buenas inferencias acerca de la población de interés.

Seleccionar un modelo razonable es a veces cuestión de actuar con base en consideraciones teóricas. Con frecuencia, por ejemplo, una situación en la que la variable aleatoria discreta de Poisson es apropiada, queda indicada por el comportamiento aleatorio de eventos en el tiempo. Sabiendo esto, podemos demostrar que el tiempo entre cualquier par adyacente de eventos sigue una distribución exponencial. Del mismo modo, si a y b son enteros, $a < b$, entonces el tiempo entre los sucesos de los eventos a-ésimo y b-ésimo posee una distribución gamma con $\alpha = b - a$. Más adelante encontraremos un teorema (llamado *teorema central del límite*) que compendia algunas condiciones que implican que una distribución normal sería una aproximación apropiada para la distribución de datos.

Una segunda forma de seleccionar un modelo es formar un histograma de frecuencia (Capítulo 1) para datos sacados de la población y escoger una función de densidad que visualmente parecería dar una curva de frecuencia similar. Por ejemplo, si un conjunto de $n = 100$ mediciones muestrales diera una distribución de frecuencia en forma de campana, podríamos concluir que la función de densidad normal representaría en forma adecuada la distribución de frecuencia poblacional.

No toda la selección de un modelo es completamente subjetiva. Hay procedimientos estadísticos para probar la hipótesis de que una distribución de frecuencia poblacional es de un tipo particular. También podemos calcular una medida de la bondad de ajuste para varias distribuciones y seleccionar la mejor. Se han hecho estudios de numerosos métodos inferenciales comunes para determinar la magnitud de los errores de inferencia introducidos por modelos incorrectos de población. Es alentador saber que muchos métodos estadísticos de inferencia son insensibles a suposiciones acerca de la forma de la distribución de frecuencia poblacional subyacente.

Las distribuciones uniforme, normal, gamma y beta ofrecen una variedad de funciones de densidad que se ajustan a numerosas distribuciones de frecuencia poblacional. Otra, la distribución Weibull, aparece en los ejercicios al final del capítulo.

4.9 Otros valores esperados

Los momentos para variables aleatorias continuas tienen definiciones análogas a las dadas para el caso discreto.

DEFINICIÓN **4.13** Si Y es una variable aleatoria continua, entonces el k–ésimo *momento alrededor del origen* está dado por

$$\mu'_k = E(Y^k), \qquad k = 1, 2, \ldots$$

El k-ésimo *momento alrededor de la media*, o el k–ésimo *momento central*, está dado por

$$\mu_k = E[(Y-\mu)^k], \qquad k = 1, 2, \ldots$$

Observe que para $k = 1$, $\mu'_1 = \mu$ y para $k = 2$, $\mu_2 = V(Y) = \sigma^2$.

EJEMPLO **4.12** Encuentre μ'_k para la variable aleatoria uniforme con $\theta_1 = 0$ y $\theta_2 = \theta$.

Solución Por definición,

$$\mu'_k = E(Y^k) = \int_{-\infty}^{\infty} y^k f(y)\, dy = \int_0^{\theta} y^k \left(\frac{1}{\theta}\right) dy = \left.\frac{y^{k+1}}{\theta(k+1)}\right]_0^{\theta} = \frac{\theta^k}{k+1}.$$

Entonces,

$$\mu'_1 = \mu = \frac{\theta}{2}, \qquad \mu'_2 = \frac{\theta^2}{3}, \qquad \mu'_3 = \frac{\theta^3}{4},$$

y así sucesivamente. ■

DEFINICIÓN **4.14** Si Y es una variable aleatoria continua, entonces la *función generadora de momento de Y* está dada por

$$m(t) = E(e^{tY}).$$

Se dice que existe la función generadora de momento si existe una constante $b > 0$ tal que $m(t)$ es finita para $|t| \le b$.

Éste es simplemente el análogo continuo de la Definición 3.14. Que $m(t)$ genere momentos está establecido en exactamente la misma forma que en la Sección 3.9. Si $m(t)$ existe, entonces

$$
\begin{aligned}
E\left(e^{tY}\right) &= \int_{-\infty}^{\infty} e^{ty} f(y)\, dy = \int_{-\infty}^{\infty} \left(1 + ty + \frac{t^2 y^2}{2!} + \frac{t^3 y^3}{3!} + \cdots\right) f(y)\, dy \\
&= \int_{-\infty}^{\infty} f(y)\, dy + t \int_{-\infty}^{\infty} y f(y)\, dy + \frac{t^2}{2!} \int_{-\infty}^{\infty} y^2 f(y)\, dy + \cdots \\
&= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots
\end{aligned}
$$

Observe que la función generadora de momento,

$$
m(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots,
$$

toma la misma forma para variables aleatorias discretas y continuas. De esta manera, el Teorema 3.12 se cumple para variables aleatorias continuas y

$$
\left.\frac{d^k m(t)}{dt^k}\right]_{t=0} = \mu_k'.
$$

EJEMPLO **4.13** Encuentre la función generadora de momento para una variable aleatoria de distribución gamma.

Solución

$$
\begin{aligned}
m(t) = E\left(e^{tY}\right) &= \int_0^{\infty} e^{ty} \left[\frac{y^{\alpha-1} e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\right] dy \\
&= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} \exp\left[-y\left(\frac{1}{\beta} - t\right)\right] dy \\
&= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} \exp\left[\frac{-y}{\beta(1-\beta t)}\right] dy.
\end{aligned}
$$

[El término exp(·) es simplemente una forma más cómoda de escribir $e^{(\cdot)}$ cuando el término del exponente es largo o complejo.]

Para completar la integración observe que la integral del factor variable de cualquier función de densidad debe ser el recíproco del factor constante. Esto es, si $f(y) = cg(y)$, donde c es una constante, entonces

$$
\int_{-\infty}^{\infty} f(y)\, dy = \int_{-\infty}^{\infty} cg(y)\, dy = 1 \text{ y por tanto} \int_{-\infty}^{\infty} g(y)\, dy = \frac{1}{c}.
$$

Aplicando este resultado a la integral en $m(t)$ y observando que si $[\beta/(1-\beta t)] > 0$ (o bien, lo que es equivalente, si $t < 1/\beta$),

$$
g(y) = y^{\alpha-1} \times \exp\{-y/[\beta/(1-\beta t)]\}
$$

es el factor variable de una función de densidad gamma con parámetros $\alpha > 0$ y $[\beta/(1-\beta t)] > 0$, obtenemos

$$
m(t) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\left[\left(\frac{\beta}{1-\beta t}\right)^{\alpha}\Gamma(\alpha)\right] = \frac{1}{(1-\beta t)^{\alpha}} \qquad \text{para } t < \frac{1}{\beta}. \qquad \blacksquare
$$

Los momentos μ'_k se pueden extraer de la función generadora de momento al derivar con respecto a t (de acuerdo con el Teorema 3.12) o al expandir la función en una serie de potencias en t. Demostraremos este último método.

EJEMPLO **4.14** Expanda la función generadora de momento del Ejemplo 4.13 en una serie de potencias en t y con ello obtenga μ'_k.

Solución Del Ejemplo 4.13, $m(t) = 1/(1 - \beta t)^{\alpha} = (1 - \beta t)^{-\alpha}$. Usando la expansión para un término binomial de la forma $(x + y)^{-c}$, tenemos

$$\begin{aligned} m(t) &= (1-\beta t)^{-\alpha} = 1 + (-\alpha)(1)^{-\alpha-1}(-\beta t) \\ &\quad + \frac{(-\alpha)(-\alpha-1)(1)^{-\alpha-2}(-\beta t)^2}{2!} + \cdots \\ &= 1 + t(\alpha\beta) + \frac{t^2[\alpha(\alpha+1)\beta^2]}{2!} + \frac{t^3[\alpha(\alpha+1)(\alpha+2)\beta^3]}{3!} + \cdots . \end{aligned}$$

Como μ'_k es el coeficiente de $t^k/k!$, encontramos, por inspección,

$$\begin{aligned} \mu'_1 &= \mu = \alpha\beta, \\ \mu'_2 &= \alpha(\alpha+1)\beta^2, \\ \mu'_3 &= \alpha(\alpha+1)(\alpha+2)\beta^3, \end{aligned}$$

y, en general, $\mu'_k = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+k-1)\beta^k$. Observe que μ'_1 y μ'_2 están acordes con los resultados del Teorema 4.8. Además, estos resultados concuerdan con el resultado del Ejercicio 4.111(a). ■

Ya hemos explicado la importancia de los valores esperados de Y^k, $(Y - \mu)^k$, y e^{tY}, todos los cuales dan información importante acerca de la distribución de Y. En ocasiones, no obstante, estamos interesados en el valor esperado de una función de una variable aleatoria como un objetivo en sí mismo. (También podemos estar interesados en la distribución de probabilidad de funciones de variables aleatorias, pero dejamos la discusión de este tema hasta el Capítulo 6.)

EJEMPLO **4.15** La energía cinética k asociada con una masa m que se mueve a una velocidad v está dada por la expresión

$$k = \frac{mv^2}{2}.$$

Considere un aparato que dispara un clavo estriado dentro de concreto a una velocidad media de 2000 pies por segundo, donde la velocidad aleatoria V posee una función de densidad dada por

$$f(v) = \frac{v^3 e^{-v/500}}{(500)^4\Gamma(4)}, \qquad v \geq 0.$$

Encuentre la energía cinética esperada asociada con un clavo de masa m.

Solución Denote con K la energía cinética aleatoria asociada con el clavo. Entonces

$$E(K) = E\left(\frac{mV^2}{2}\right) = \frac{m}{2}E(V^2),$$

por el Teorema 4.5, parte 2. La variable aleatoria V tiene una distribución gamma con $\alpha = 4$ y $\beta = 500$. Por tanto, $E(V^2) = \mu_2'$ para la variable aleatoria V. Por consulta del Ejemplo 4.14, tenemos $\mu_2' = \alpha(\alpha+1)\beta^2 = 4(5)(500)^2 = 5{,}000{,}000$. En consecuencia,

$$E(K) = \frac{m}{2}E(V^2) = \frac{m}{2}(5{,}000{,}000) = 2{,}500{,}000\ m. \qquad \blacksquare$$

Hallar los momentos de una función de una variable aleatoria con frecuencia se facilita con el uso de su función generadora de momento.

TEOREMA 4.12

Sea Y una variable aleatoria con función de densidad $f(y)$ y $g(Y)$ una función de Y. Entonces la función generadora de momento para $g(Y)$ es

$$E[e^{tg(Y)}] = \int_{-\infty}^{\infty} e^{tg(y)} f(y)\, dy.$$

Este teorema se concluye directamente de la Definición 4.14 y el Teorema 4.4.

EJEMPLO 4.16 Sea $g(\overline{Y}) = Y - \mu$, donde Y es una variable aleatoria normalmente distribuida con media μ y varianza σ^2. Encuentre la función generadora de momento para $g(Y)$.

Solución La función generadora de momento de $g(Y)$ está dada por

$$m(t) = E[e^{tg(Y)}] = E[e^{t(Y-\mu)}] = \int_{-\infty}^{\infty} e^{t(y-\mu)}\left[\frac{\exp[-(y-\mu)^2/2\sigma^2]}{\sigma\sqrt{2\pi}}\right] dy.$$

Para integrar, sea $u = y - \mu$. Entonces $du = dy$ y

$$\begin{aligned} m(t) &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tu}e^{-u^2/(2\sigma^2)}\, du \\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(u^2 - 2\sigma^2 tu)\right] du. \end{aligned}$$

Complete el cuadrado del exponente de e al multiplicar y dividir entre $e^{t^2 s^{2}/2}$. Entonces

$$\begin{aligned} m(t) &= e^{t^2\sigma^2/2}\int_{-\infty}^{\infty} \frac{\exp[-(1/2\sigma^2)(u^2 - 2\sigma^2 tu + \sigma^4 t^2)]}{\sigma\sqrt{2\pi}}\, du \\ &= e^{t^2\sigma^2/2}\int_{-\infty}^{\infty} \frac{\exp[-(u-\sigma^2 t)^2\ \ 2\sigma^2]}{\sigma\sqrt{2\pi}}\, du. \end{aligned}$$

La función dentro de la integral es una función de densidad con media $\sigma^2 t$ y varianza σ^2. (Vea la ecuación para la función de densidad normal en la Sección 4.5.) Por tanto, la integral es igual a 1 y

$$m(t) = e^{(t^2/2)\sigma^2}.$$

Los momentos de $U = Y - \mu$ se pueden obtener de $m(t)$ al derivar $m(t)$ de acuerdo con el Teorema 3.12 o al expandir $m(t)$ en una serie. ■

El propósito del análisis de momentos anterior es doble. Primero, se pueden usar momentos como medidas descriptivas numéricas para explicar los datos que obtenemos en un experimento. En segundo término, se pueden usar en un sentido teórico para demostrar que una variable aleatoria posee una distribución de probabilidad particular. Se puede demostrar que si dos variables aleatorias Y y Z poseen funciones generadoras de momento idénticas, entonces Y y Z poseen distribuciones de probabilidad idénticas. Esta última aplicación de los momentos se mencionó en el análisis de las funciones generadoras de momento para variables aleatorias discretas en la Sección 3.9; también se aplica a variables aleatorias continuas.

Para comodidad del estudiante, las funciones de probabilidad y densidad, medias, varianzas y funciones generadoras de momento para algunas variables aleatorias comunes se dan en el Apéndice 2 y al final de este libro.

Ejercicios

4.136 Suponga que el tiempo de espera para que el primer cliente entre en una tienda de venta al menudeo, después de las 9:00 a.m., es una variable aleatoria Y con una función de densidad exponencial dada por

$$f(y) = \begin{cases} \left(\frac{1}{\theta}\right) e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la función generadora de momento para Y.

b Use la respuesta del inciso a para hallar $E(Y)$ y $V(Y)$.

4.137 Demuestre que el resultado dado en el Ejercicio 3.158 también se cumple para variables aleatorias continuas. Esto es, demuestre que si Y es una variable aleatoria con función generadora de momento $m(t)$ y U está dada por $U = aY + b$, la función generadora de momento de U es $e^{tb}m(at)$. Si Y tiene media μ y varianza σ^2 use la función generadora de momento de U para obtener la media y la varianza de U.

4.138 En el Ejemplo 4.16 se obtuvo la función generadora de momento para $Y - \mu$, donde Y está distribuida normalmente con media μ y varianza σ^2.

a Use los resultados del Ejemplo 4.16 y el Ejercicio 4.137 para hallar la función generadora de momento para Y.

b Obtenga la derivada de la función generadora de momento hallada en el inciso a para demostrar que $E(Y) = \mu$ y $V(Y) = \sigma^2$.

4.139 En el Ejercicio 4.138 se demostró que la función generadora de momento de una variable aleatoria normalmente distribuida, Y, con media μ y varianza σ^2 es $m(t) = e^{\mu t + (1/2)t^2\sigma^2}$. Use el resultado del Ejercicio 4.137 para derivar la función generadora de momento de $X = -3Y + 4$. ¿Cuál es la distribución de X? ¿Por qué?

4.140 Identifique las distribuciones de las variables aleatorias con las siguientes funciones generadoras de momento:

a $m(t) = (1 - 4t)^{-2}$.

b $m(t) = 1(\ 1 - 3.2t)$.

c $m(t) = e^{-5t+6t^2}$.

4.141 Si $\theta_1 < \theta_2$, obtenga la función generadora de momento de una variable aleatoria que tenga una distribución uniforme en el intervalo (θ_1, θ_2).

4.142 Consulte los Ejercicios 4.141 y 4.137. Suponga que Y está uniformemente distribuida en el intervalo (0, 1) y que $a > 0$ es una constante.

a Obtenga la función generadora de momento para Y.

b Obtenga la función generadora de momento de $W = aY$. ¿Cuál es la distribución de W? ¿Por qué?

c Obtenga la función generadora de momento de $X = -aY$. ¿Cuál es la distribución de X? ¿Por qué?

d Si b es una constante fija, obtenga la función generadora de momento de $V = aY + b$. ¿Cuál es la distribución de V? ¿Por qué?

4.143 La función generadora de momento para la variable aleatoria gamma se obtuvo en el Ejemplo 4.13. Encuentre la derivada de esta función generadora de momento para hallar la media y la varianza de la distribución gamma.

4.144 Considere una variable aleatoria Y con función de densidad dada por

$$f(y) = ke^{-y^2/2}, \qquad -\infty < y < \infty.$$

a Encuentre k.

b Encuentre la función generadora de momento de Y.

c Encuentre $E(Y)$ y $V(Y)$.

4.145 Una variable aleatoria Y tiene la función de densidad

$$f(y) = \begin{cases} e^y, & y < 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $E(e^{3Y/2})$.

b Encuentre la función generadora de momento para Y.

c Encuentre $V(Y)$.

4.10 Teorema de Tchebysheff

Como fue el caso para variables aleatorias discretas, una interpretación de μ y σ para variables aleatorias continuas está dada por la regla empírica y el teorema de Tchebysheff. Incluso si las distribuciones exactas son desconocidas para variables aleatorias de interés, el conocimiento de las medias y desviaciones estándar asociadas nos permiten deducir límites significativos para las probabilidades de eventos que con frecuencia son de interés.

Expresamos y utilizamos el teorema de Tchebysheff en la Sección 3.11. Ahora volvemos a expresar este teorema y damos una prueba aplicable a una variable aleatoria continua.

TEOREMA **4.13**

Teorema de Tchebysheff Sea Y una variable aleatoria con media finita μ y varianza σ^2. Entonces, para cualquier $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{o} \quad P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Demostración Daremos la demostración para una variable aleatoria continua. La prueba para el caso discreto se desarrolla de un modo similar. Denote con $f(y)$ la función de densidad de Y. Entonces,

$$\begin{aligned} V(Y) = \sigma^2 &= \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\, dy \\ &= \int_{-\infty}^{\mu-k\sigma} (y-\mu)^2 f(y)\, dy + \int_{\mu-k\sigma}^{\mu+k\sigma} (y-\mu)^2 f(y)\, dy \\ &\quad + \int_{\mu+k\sigma}^{\infty} (y-\mu)^2 f(y)\, dy. \end{aligned}$$

La segunda integral es siempre mayor o igual a cero y $(y-\mu)^2 \geq k^2\sigma^2$ para todos los valores de y entre los límites de integración para las integrales primera y tercera; esto es, las regiones de integración están en las colas de la función de densidad e incluyen sólo valores de y para los cuales $(y-\mu)^2 \geq k^2\sigma^2$. Sustituya con cero la segunda integral y sustituya $k^2\sigma^2$ por $(y-\mu)^2$ en las integrales primera y tercera para obtener la desigualdad

$$V(Y) = \sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(y)\, dy + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(y)\, dy.$$

Entonces

$$\sigma^2 \geq k^2\sigma^2 \left[\int_{-\infty}^{\mu-k\sigma} f(y)\, dy + \int_{\mu+k\sigma}^{+\infty} f(y)\, dy \right],$$

o bien

$$\sigma^2 \geq k^2\sigma^2 [P(Y \leq \mu - k\sigma) + P(Y \geq \mu + k\sigma)] = k^2\sigma^2 P(|Y-\mu| \geq k\sigma).$$

Dividiendo entre $k^2\sigma^2$, obtenemos

$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2},$$

o bien, lo que es equivalente,

$$P(\ Y-\mu\ < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Un valor real del teorema de Tchebysheff es que hace posible que encontremos límites para probabilidades que de ordinario tendrían que obtenerse por tediosas manipulaciones matemáticas (integración o sumatoria). Además, con frecuencia podemos obtener medias y varianzas de variables aleatorias (vea el Ejemplo 4.15) sin especificar la *distribución* de la variable. En situaciones como esta, el teorema de Tchebysheff todavía proporciona límites significativos para probabilidades de interés.

EJEMPLO **4.17** Suponga que la experiencia ha demostrado que el tiempo Y (en minutos) necesario para realizar una prueba periódica de mantenimiento en una máquina de dictados sigue una distribución gamma con $\alpha = 3.1$ y $\beta = 2$. Un trabajador de mantenimiento de nuevo ingreso tarda 22.5

minutos en probar la máquina. ¿El tiempo para realizar la prueba está en desacuerdo con la experiencia anterior?

Solución La media y la varianza para los tiempos de prueba de mantenimiento (basados en la experiencia anterior) son (del Teorema 4.8)

$$\mu = \alpha\beta = (3.1)(2) = 6.2 \quad \text{y} \quad \sigma^2 = \alpha\beta^2 = (3.1)(2^2) = 12.4.$$

Se deduce que $\sigma = \sqrt{12.4} = 3.52$. Observe que $y = 22.5$ minutos es mayor que la media de $\mu = 6.2$ minutos por 16.3 minutos o $k = 16.3/3.52 = 4.63$ desviaciones estándar. Entonces, del teorema de Tchebysheff,

$$P(|Y - 6.2| \geq 16.3) = P(|Y - \mu| \geq 4.63\sigma) \leq \frac{1}{(4.63)^2} = .0466$$

Esta probabilidad está basada en la suposición de que la distribución de tiempos de mantenimiento no ha cambiado con respecto a la experiencia anterior. Entonces, observando que $P(Y \geq 22.5)$ es pequeña, debemos concluir que nuestro trabajador de nuevo ingreso ha generado por casualidad un largo tiempo de mantenimiento, lo cual sucede con poca probabilidad, o bien que el nuevo trabajador es más lento que los anteriores. Considerando la baja probabilidad de $P(Y \geq 22.5)$, estamos a favor de este último punto de vista. ■

La probabilidad exacta, $P(Y \geq 22.5)$, para el Ejemplo 4.17 requeriría evaluar la integral

$$P(Y \geq 22.5) = \int_{22.5}^{\infty} \frac{y^{2.1}e^{-y/2}}{2^{3.1}\Gamma(3.1)}\, dy.$$

Aun cuando podríamos utilizar tablas dadas por Pearson (1965) para evaluar esta integral, no podemos hacerlo directamente. Por supuesto que podríamos usar *R* o *S*–Plus o una de las aplicaciones breves para evaluar numéricamente esta probabilidad. A menos que usemos software de estadística, integrales similares son difíciles de evaluar para la densidad beta y para muchas otras funciones de densidad. El teorema de Tchebysheff a veces proporciona límites rápidos para probabilidades a la vez que evita una laboriosa integración, la utilización de software o búsquedas de tablas apropiadas.

Ejercicios

4.146 Un fabricante de llantas desea calcular un intervalo de rendimiento en millas que excluya no más de 10% del rendimiento de las llantas que él vende. Todo lo que sabe es que, para un gran número de llantas probadas, la media de rendimiento fue de 25,000 millas y que la desviación estándar fue de 4000 millas. ¿Qué intervalo sugeriría usted?

4.147 Una máquina empleada para llenar cajas de cereal despacha, en promedio, μ onzas por caja. El fabricante desea que las Y onzas reales despachadas no rebasen por más de 1 onza a μ, al menos 75% del tiempo. ¿Cuál es el máximo valor de σ, la desviación estándar de Y, que se puede tolerar si las metas del fabricante han de satisfacerse?

4.148 Encuentre $P(|Y - \mu| \leq 2\sigma)$ para el Ejercicio 4.16. Compare con los correspondientes enunciados probabilísticos dados por el teorema de Tchebysheff y la regla empírica.

4.149 Determine $P(|Y - \mu| \leq 2\sigma)$ para la variable aleatoria uniforme. Compare con los correspondientes enunciados probabilísticos dados por el teorema de Tchebysheff y la regla empírica.

4.150 Determine $P(|Y - \mu| \leq 2\sigma)$ para la variable aleatoria exponencial. Compare con los correspondientes enunciados probabilísticos dados por el teorema de Tchebysheff y la regla empírica.

4.151 Consulte el Ejercicio 4.92. ¿Esperaría que C excediera de 2000 con frecuencia?

4.152 Consulte el Ejercicio 4.109. Encuentre un intervalo que contenga L durante al menos 89% de las semanas que la máquina está en uso.

4.153 Consulte el Ejercicio 4.129. Encuentre un intervalo para el cual la probabilidad de que C se encuentre dentro del mismo sea al menos de .75.

4.154 Suponga que Y es una variable aleatoria distribuida con χ^2, con $\nu = 7$ grados de libertad.

a ¿Cuáles son la media y la varianza de Y?

b ¿Es probable que Y tome un valor de 23 o más?

c **Ejercicio Applet** Use la aplicación *Gamma Probabilities and Quantiles* para hallar $P(Y > 23)$.

4.11 Valores esperados de funciones discontinuas y distribuciones mixtas de probabilidad (opcional)

Los problemas en probabilidad y estadística en ocasiones incluyen funciones que son parcialmente continuas y parcialmente discretas, en una de dos maneras. Primero, podemos estar interesados en las propiedades, por ejemplo el valor esperado de una variable aleatoria $g(Y)$ que es una función discontinua de una variable aleatoria discreta o continua Y. En segundo término, la variable aleatoria de interés misma puede tener una función de distribución que es continua sobre algunos intervalos de manera que algunos puntos aislados tienen probabilidades positivas.

Ilustramos estas ideas con los siguientes ejemplos.

EJEMPLO **4.18** Un vendedor minorista de un producto derivado del petróleo vende una cantidad aleatoria Y al día. Suponga que Y, medida en miles de galones, tiene la función de densidad de probabilidad

$$f(y) = \begin{cases} (3/8)y^2, & 0 \leq y \leq 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La utilidad del minorista resulta ser de \$100 por cada 1000 galones vendidos (10¢ por galón) si $Y \leq 1$ y \$40 *extra* por 1000 galones (4¢ extra por galón) si $Y > 1$. Encuentre la utilidad esperada del minorista para cualquier día determinado.

Solución Denote con $g(Y)$ la utilidad diaria del minorista. Entonces

$$g(Y) = \begin{cases} 100Y, & 0 \leq Y \leq 1, \\ 140Y, & 1 < Y \leq 2. \end{cases}$$

Deseamos hallar la utilidad esperada; por el Teorema 4.4, el valor esperado es

$$
\begin{aligned}
E[g(Y)] &= \int_{-\infty}^{\infty} g(y)f(y)\,dy \\
&= \int_0^1 100y\left[\left(\frac{3}{8}\right)y^2\right] dy + \int_1^2 140y\left[\left(\frac{3}{8}\right)y^2\right] dy \\
&= \frac{300}{(8)(4)}y^4\Big]_0^1 + \frac{420}{(8)(4)}y^4\Big]_1^2 \\
&= \frac{300}{32}(1) + \frac{420}{32}(15) = 206.25.
\end{aligned}
$$

Entonces, el minorista puede esperar una utilidad de \$206.25 con la venta diaria de este producto particular. ■

Suponga que Y denota la cantidad que paga por póliza en un año una compañía de seguros que proporciona seguro a automóviles. Para muchas pólizas, $Y = 0$ porque las personas aseguradas no están involucradas en accidentes. Para personas aseguradas que *sufren* accidentes, la cantidad pagada por la compañía podría ser modelada con una de las funciones de densidad que previamente hemos estudiado. Una variable aleatoria Y que tiene alguna de sus probabilidades en puntos discretos (0 en este ejemplo) y el resto disperso en intervalos, se dice que tiene una *distribución mezclada*. Denote con $F(y)$ una función de distribución de una variable aleatoria Y que tiene una distribución mezclada. Para todos los fines prácticos, cualquier función $F(y)$ de distribución mezclada se puede escribir de manera única como

$$F(y) = c_1F_1(y) + c_2F_2(y),$$

donde $F_1(y)$ es una función escalón de distribución, $F_2(y)$ es una función de distribución continua, c_1 es la probabilidad acumulada de todos los puntos discretos y $c_2 = 1 - c_1$ es la probabilidad acumulada de todas las porciones continuas.

El siguiente ejemplo da una ilustración de una distribución mezclada.

EJEMPLO **4.19** Denote con Y la vida útil (en cientos de horas) de componentes electrónicos. Éstos fallan con frecuencia inmediatamente despúes de conectarlos en un sistema. Se ha observado que la probabilidad de falla inmediata es $1/4$. Si un componente no falla de inmediato, la distribución para su vida útil tiene la función de densidad exponencial

$$f(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de distribución para Y y evalúe $P(Y > 10)$.

Solución Hay sólo un punto discreto, $y = 0$, y este punto tiene probabilidad $1/4$. Por tanto, $c_1 = 1/4$ y $c_2 = 3/4$. Se deduce que Y es una mezcla de las distribuciones de dos variables aleatorias, X_1

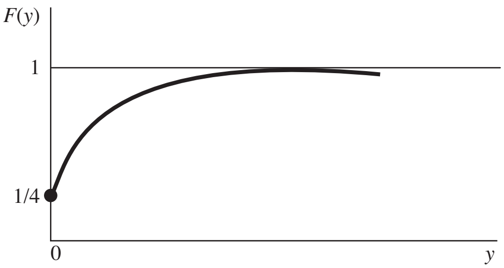


FIGURA **4.18** Función de distribución $F(y)$ para el Ejemplo 4.19

y X_2, donde X_1 tiene probabilidad 1 en el punto 0 y X_2 tiene la densidad exponencial dada. Esto es,

$$F_1(y) = \begin{cases} 0, & y < 0, \\ 1, & y \geq 0, \end{cases}$$

y

$$F_2(y) = \begin{cases} 0, & y < 0, \\ \int_0^y e^{-x} dx = 1 - e^{-y}, & y \geq 0. \end{cases}$$

Ahora

$$F(y) = (1/4)F_1(y) + (3/4)F_2(y),$$

y, por tanto,

$$\begin{aligned} P(Y > 10) &= 1 - P(Y \leq 10) = 1 - F(10) \\ &= 1 - [(1/4) + (3/4)(1 - e^{-10})] \\ &= (3/4)[1 - (1 - e^{-10})] = (3/4)e^{-10}. \end{aligned}$$

En la Figura 4.18 se da una gráfica de $F(y)$. ■

Un método fácil para hallar el valor esperado de variables aleatorias con distribuciones mezcladas se da en la Definición 4.15.

DEFINICIÓN **4.15**

Hagamos que Y tenga la función de distribución mezclada

$$F(y) = c_1 F_1(y) + c_2 F_2(y)$$

y supongamos que X_1 es una variable aleatoria discreta con función de distribución $F_1(y)$ y que X_2 es una variable aleatoria continua con función de distribución $F_2(y)$. Denotemos con $g(Y)$ una función de Y. Entonces

$$E[g(Y)] = c_1 E[g(X_1)] + c_2 E[g(X_2)].$$

EJEMPLO **4.20** Encuentre la media y la varianza de la variable aleatoria definida en el Ejemplo 4.19.

Solución Con todas las definiciones como en el Ejemplo 4.19, se deduce que

$$E(X_1) = 0 \quad \text{y} \quad E(X_2) = \int_0^\infty ye^{-y}\,dy = 1.$$

Por tanto,

$$\mu = E(Y) = (1/4)E(X_1) + (3/4)E(X_2) = 3/4.$$

También,

$$E(X_1^2) = 0 \quad \text{y} \quad E(X_2^2) = \int_0^\infty y^2e^{-y}\,dy = 2.$$

En consecuencia,

$$E(Y^2) = (1/4)E(X_1^2) + (3/4)E(X_2^2) = (1/4)(0) + (3/4)(2) = 3/2.$$

Entonces

$$V(Y) = E(Y^2) - \mu^2 = (3/2) - (3/4)^2 = 15/16.$$

■

Ejercicios

***4.155** Una constructora de casas necesita adquirir algunos abastecimientos que tienen un tiempo de espera Y para entregarse, con una distribución uniforme continua en el intervalo de 1 a 4 días. Como ella puede arreglárselas sin el material durante 2 días, el costo de la demora se fija en \$100 para cualquier tiempo de espera hasta de 2 días. Después de 2 días, no obstante, el costo de la demora es \$100 más \$20 por día (prorrateado) por cada día individual. Esto es, si el tiempo de espera es 3.5 días, el costo de la demora es \$100 + \$20(1.5) = \$130. Encuentre el valor esperado del costo de la constructora debido a la espera de abastecimientos.

***4.156** La duración Y de llamadas telefónicas de larga distancia (en minutos) vigilada por una estación es una variable aleatoria con propiedades tales que

$$P(Y = 3) = .2 \quad \text{y} \quad P(Y = 6) = .1.$$

De otro modo, Y tiene una función de densidad continua dada por

$$f(y) = \begin{cases} (1/4)ye^{-y/2}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Los puntos discretos en 3 y 6 se deben al hecho de que la duración de la llamada se anuncia a quien genera la llamada en intervalos de tres minutos, y quien llama debe pagar tres minutos incluso si habla durante un periodo menor. Encuentre la duración esperada de una llamada de larga distancia seleccionada al azar.

***4.157** Se sabe que la vida útil Y de un componente empleado en un complejo sistema electrónico tiene una densidad exponencial con una media de 100 horas. El componente se cambia cuando falla o a una edad de 200 horas, lo que ocurra primero.

a Encuentre la función de distribución para X, la vida útil en que el componente está en uso.

b Encuentre $E(X)$.

*4.158 Considere el aparato que dispara clavos del Ejemplo 4.15. Cuando el aparato funciona, el clavo es disparado con velocidad, V, con densidad

$$f(v) = \frac{v^3 e^{-v/500}}{(500)^4 \Gamma(4)}.$$

El aparato no dispara bien 2% del tiempo que se usa, resultando en una velocidad de 0. Encuentre la energía cinética esperada asociada con un clavo de masa m. Recuerde que la energía cinética, k, de una masa m que se mueve a velocidad v es $k = (mv^2)/2$.

*4.159 Una variable aleatoria Y tiene función de distribución

$$F(y) = \begin{cases} 0, & \text{si } y < 0, \\ y^2 + 0.1, & \text{si } 0 \le y < .5, \\ y, & \text{si } .5 \le y < 1, \\ 1, & \text{si } y \ge 1. \end{cases}$$

a Dé $F_1(y)$ y $F_2(y)$, los componentes discretos y continuos de $F(y)$.

b Escriba $F(y)$ como $c_1F_1(y) + c_2F_2(y)$.

c Encuentre el valor y la varianza esperados de Y.

4.12 Resumen

Este capítulo presentó modelos probabilísticos para variables aleatorias continuas. La función de densidad, que da un modelo para una distribución de frecuencia poblacional asociada con una variable aleatoria continua, de manera subsiguiente dará un mecanismo para inferir características de la población basada en mediciones contenidas en una muestra tomada de esa población. Como consecuencia, la función de densidad da un modelo para una distribución real de datos que existe o podría ser generada por experimentación repetida. Distribuciones similares para pequeños conjuntos de datos (muestras de poblaciones) se estudiaron en el Capítulo 1.

Cuatro tipos específicos de funciones de densidad: uniforme, normal, gamma (con la χ^2 y exponencial como casos especiales) y beta se presentaron y dan una amplia variedad de modelos para distribuciones de frecuencia poblacional. Para comodidad del estudiante, la Tabla 4.1 contiene un resumen de los comandos *R* (o *S*-Plus) que dan probabilidades y cuantiles asociados con estas distribuciones. Muchas otras funciones de densidad podrían utilizarse para ajustarse a situaciones reales, pero las cuatro descritas se adaptan perfectamente en muchas situaciones. Otras funciones de densidad más se presentan en los ejercicios al final del capítulo.

Lo adecuado de una función de densidad para modelar la distribución de frecuencia para una variable aleatoria depende de la técnica para hacer inferencias que se emplee. Si un modesto desa-

Tabla 4.1 Procedimientos *R* (y *S*-Plus) que dan probabilidades y percentiles para algunas distribuciones continuas comunes

Distribución	$P(Y \le y_0)$	p–ésimo cuantil: ϕ_p manera que $P(Y \le \phi_p) = p$
Normal	pnorm(y_0,μ,σ)	qnorm(p,μ,σ)
Exponencial	pexp(y_0,1/β)	qexp(p,1/β)
Gamma	pgamma(y_0,α,1/β)	qgamma(p,α,1/β)
Beta	pbeta(y_0,α,β)	qbeta(p,α,β)

cuerdo entre el modelo y la distribución de frecuencia de población real no afecta la bondad del procedimiento inferencial, el modelo es adecuado.

La última parte del capítulo se ocupó de valores esperados, en particular momentos y funciones generadoras de momento. Es importante concentrar la atención en la razón para presentar estas cantidades y evitar excesiva concentración en los aspectos matemáticos del material. Los momentos, particularmente la media y la varianza, son medidas descriptivas numéricas para variables aleatorias. En especial, más adelante veremos que a veces es difícil hallar la distribución de probabilidad para una variable aleatoria Y o una función $g(Y)$, y ya hemos observado que la integración en intervalos para muchas funciones de densidad (la normal y gamma, por ejemplo) es muy difícil. Cuando esto ocurre, podemos describir de un modo aproximado el comportamiento de la variable aleatoria con el uso de sus momentos, junto con el teorema de Tchebysheff y la regla empírica (Capítulo 1).

Bibliografía y lecturas adicionales

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1995. *Continuous Univariate Distributions*, 2d ed. New York: Wiley.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Pearson, K., ed. 1965. *Tables of the Incomplete Gamma Function*. London: Cambridge University Press.

———. 1968. *Tables of the Incomplete Beta Function*. London: Cambridge University Press.

Perruzzi, J. J., and E. J. Hilliard. 1984. "Modeling Time-Delay Measurement Errors Using a Generalized Beta Density Function," *Journal of the Acoustical Society of America* 75(1): 197–201.

Tables of the Binomial Probability Distribution. 1950. Department of Commerce, National Bureau of Standards, Applied Mathematics Series 6.

Zimmels, Y. 1983. "Theory of Kindered Sedimentation of Polydisperse Mixtures," *American Institute of Chemical Engineers Journal* 29(4): 669–76.

Zwilliger, D. 2002. *CRC Standard Mathematical Tables*, 31st ed. Boca Raton, Fla.: CRC Press.

Ejercicios complementarios

4.160 Sea la función de densidad de una variable aleatoria Y dada por

$$f(y) = \begin{cases} \dfrac{2}{\pi(1+y^2)}, & -1 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la función de distribución.

b Encuentre $E(Y)$.

4.161 El tiempo necesario para completar un examen de aptitud en universidades se encuentra normalmente distribuido con media de 70 minutos y desviación estándar de 12 minutos. ¿Cuánto debe durar el examen si deseamos que 90% de los estudiantes tenga suficiente tiempo para completar el examen?

4.162 Una fábrica utiliza 3000 focos cuya vida útil está normalmente distribuida con media de 500 horas y desviación estándar de 50. Para reducir al mínimo el número de focos que se queman durante horas de operación, todos son cambiados después de un periodo determinado. ¿Con qué frecuencia deben cambiarse los focos si deseamos que no más de 1% de los focos se quemen entre periodos de cambio?

4.163 Consulte el Ejercicio 4.66. Suponga que cinco cojinetes se sacan al azar de la producción. ¿Cuál es la probabilidad de que al menos uno esté defectuoso?

4.164 La vida útil de las barrenas de perforación de pozos petroleros depende de los tipos de roca y suelo que encuentren al perforar, pero se estima que la duración media es de 75 horas. Una compañía de exploración compra barrenas cuya vida útil está normalmente distribuida, en forma aproximada, con media de 75 horas y desviación estándar de 12 horas. ¿Qué proporción de las barrenas de la compañía

a fallarán antes de 60 horas de uso?,

b durarán al menos 60 horas?,

c tendrán que ser cambiadas después de más de 90 horas de uso?

4.165 Sea que Y tenga función de densidad

$$f(y) = \begin{cases} cye^{-2y}, & 0 \leq y \leq \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de c que haga de $f(y)$ una función de densidad.

b Obtenga la media y la varianza para Y.

c Obtenga la función generadora de momento para Y.

4.166 Use el hecho de que

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots$$

para expandir la función generadora de momento del Ejemplo 4.16 en una serie para hallar μ_1, μ_2, μ_3 y μ_4 para la variable aleatoria normal.

4.167 Encuentre una expresión para $\mu'_k = E(Y^k)$, donde la variable aleatoria Y tiene una distribución beta.

4.168 El número de llegadas N a una caja de un supermercado en el intervalo de 0 a t sigue una distribución de Poisson con media λt. Denote con T el tiempo hasta la primera llegada. Encuentre la función de densidad para T. [*Nota:* $P(T > t_0) = P(N = 0 \text{ en } t = t_0)$.]

4.169 Se puede usar un argumento semejante al del Ejercicio 4.168 para demostrar que si los eventos ocurren en el tiempo de acuerdo con una distribución de Poisson con media λt, entonces los tiempos de llegada entre eventos tienen una distribución exponencial con media $1/\lambda$. Si entran llamadas a un centro policial de emergencia a razón de diez por hora, ¿cuál es la probabilidad de que transcurran más de 15 minutos entre las dos llamadas siguientes?

***4.170** Consulte el Ejercicio 4.168.

a Si U es el tiempo hasta la *segunda* llegada, demuestre que U tiene una función de densidad gamma con $\alpha = 2$ y $\beta = 1/\lambda$.

b Demuestre que el tiempo hasta la k–ésima llegada tiene una densidad gamma con $\alpha = k$ y $\beta = 1/\lambda$.

4.171 Suponga que llegan clientes a una caja a razón de dos por minuto.

a ¿Cuáles son la media y la varianza de los tiempos de espera entre llegadas sucesivas de clientes?

b Si una empleada tarda tres minutos en atender al primer cliente que llega a una caja, ¿cuál es la probabilidad de que al menos un cliente más esté esperando cuando se termine de servir al primer cliente?

4.172 Las llamadas para conexiones de entrada a un centro de computación llegan a un ritmo promedio de cuatro por minuto. Las llamadas siguen una distribución de Poisson. Si llega una llamada al principio de un intervalo de un minuto, ¿cuál es la probabilidad de que una segunda llamada no llegue en los siguientes 20 segundos?

4.173 Suponga que plantas de una especie particular se dispersan en una superficie de modo que el número de ellas en un área determinada sigue una distribución de Poisson con densidad media de λ plantas por unidad de área. Si una planta se selecciona al azar en esta área, encuentre la función de densidad de probabilidad de la distancia a la planta vecina *más cercana*. [*Sugerencia:* Si R denota la distancia a la vecina más próxima, entonces $P(R > r)$ es igual a la probabilidad de no ver plantas en un círculo de radio r.]

4.174 El tiempo (en horas) que un gerente tarda en entrevistar a un solicitante de trabajo tiene una distribución exponencial con $\beta = 1/2$. Los solicitantes son citados a intervalos de un cuarto de hora, empezando a las 8:00 a.m. y llegan exactamente a tiempo. Cuando el candidato citado a las 8:15 a.m. llega a la oficina del gerente, ¿cuál es la probabilidad de que tenga que esperar antes de ser entrevistado?

4.175 El valor de la mediana y de una variable aleatoria continua es aquel para el que $F(y) = .5$. Encuentre el valor de la mediana de la variable aleatoria del Ejercicio 4.11.

4.176 Si Y tiene una distribución exponencial con media β, encuentre (como función de β) la mediana de Y.

4.177 **Ejercicio Applet** Use la aplicación *Gamma Probabilities and Quantiles* para hallar las medianas de variables aleatorias con distribución gamma y parámetros

a $\alpha = 1, \beta = 3$. Compare su respuesta con la del Ejercicio 4.176.

b $\alpha = 2, \beta = 2$. ¿La mediana es mayor o menor que $E(Y)$?

c $\alpha = 5, \beta = 10$. ¿La mediana es mayor o menor que $E(Y)$?

d En todos estos casos, la mediana excede a la media. ¿Cómo se refleja eso en las formas de las densidades correspondientes?

4.178 Grafique la función de densidad de probabilidad beta para $\alpha = 3$ y $\beta = 2$.

a Si Y tiene esta función de densidad beta, encuentre $P(.1 \le Y \le .2)$ usando probabilidades binomiales para evaluar $F(y)$. (Vea la Sección 4.7.)

b **Ejercicio Applet** Si Y tiene esta función de densidad beta, encuentre $P(.1 \le Y \le .2)$, usando la aplicación *Beta Probabilities and Quantiles.*

c **Ejercicio Applet** Si Y tiene esta función de densidad beta, use la aplicación *Beta Probabilities and Quantiles* para hallar los cuantiles .05 y .95 para Y.

d ¿Cuál es la probabilidad de que Y caiga entre los dos cuantiles encontrados en el inciso c?

***4.179** Una comerciante al menudeo tiene una demanda diaria Y de cierto alimento que se vende por libra, donde Y (medido en cientos de libras) tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(Ella no puede tener en existencia más de 100 libras.) La comerciante desea adquirir $100k$ libras de alimento, que compra en 6¢ por libra y lo vende en 10¢ por libra. ¿Qué valor de k minimizará su utilidad diaria esperada?

4.180 Suponga que Y tiene una distribución gamma con $\alpha = 3$ y $\beta = 1$.

a Use probabilidades de Poisson para evaluar $P(Y \le 4)$. (Vea el Ejercicio 4.99.)

b **Ejercicio Applet** Use la aplicación *Gamma Probabilities and Quantiles* para hallar $P(Y \le 4)$.

4.181 Suponga que Y es una variable aleatoria normalmente distribuida con media μ y varianza σ^2. Use los resultados del Ejemplo 4.16 para hallar la función generadora de momento, media y varianza de

$$Z = \frac{Y - \mu}{\sigma}.$$

¿Cuál es la distribución de Z? ¿Por qué?

***4.182** Se dice que una variable aleatoria Y tiene una distribución log–normal si $X = \ln(Y)$ tiene una distribución normal. (El símbolo ln denota logaritmo natural.) En ese caso Y debe ser no negativa. La forma de la función de densidad de probabilidad log-normal es semejante a la de la distribución gamma, con una larga cola a la derecha. La ecuación de la función de densidad log-normal está dada por

$$f(y) = \begin{cases} \dfrac{1}{\sigma y\sqrt{2\pi}} e^{-(\ln(y)-\mu)^2(\ 2\sigma^2)}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Como $\ln(y)$ es una función monotónica de y,

$$P(Y \le y) = P[\ln(Y) \le \ln(y)] = P[X \le \ln(y)],$$

donde X tiene una distribución normal con media μ y varianza σ^2. De esta forma, las probabilidades para variables aleatorias con una distribución log-normal se pueden hallar al transformarlas en probabilidades que puedan calcularse usando la distribución normal ordinaria. Si Y tiene una distribución log-normal con $\mu = 4$ y $\sigma^2 = 1$, encuentre

a $P(Y \le 4)$.

b $P(Y > 8)$.

4.183 Si Y tiene una distribución log-normal con parámetros μ y σ^2 se puede demostrar que

$$E(Y) = e^{(\mu+\sigma^2)/2} \quad \text{y} \quad V(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).$$

Los granos que componen metales policristalinos tienden a tener pesos que siguen una distribución log-normal. Para un tipo de aluminio, los pesos en gramos tienen una distribución log-normal con $\mu = 3$ y $\sigma = 4$ (en unidades de 10^{-2} g).

a Encuentre la media y la varianza de los pesos en los granos.

b Encuentre un intervalo en el que al menos 75% de los pesos en granos deben estar. [*Sugerencia:* Use el teorema de Tchebysheff.]

c Encuentre la probabilidad de que un grano seleccionado al azar pese menos que el peso medio del grano.

4.184 Denote con Y una variable aleatoria con función de densidad de probabilidad dada por

$$f(y) = (1/2)e^{-|y|}, \qquad -\infty < y < \infty.$$

Encuentre la función generadora de momento de Y y úsela para hallar $E(Y)$.

***4.185** Sean $f_1(y)$ y $f_2(y)$ funciones de densidad y sea a una constante tal que $0 \le a \le 1$. Considere la función $f(y) = af_1(y) + (1 - a)f_2(y)$.

a Demuestre que $f(y)$ es una función de densidad que se conoce con frecuencia como mezcla de dos funciones de densidad.

b Suponga que Y_1 es una variable aleatoria con función de densidad $f_1(y)$ y que $E(Y_1) = \mu_1$ y $\text{Var}(Y_1) = \sigma_1^2$; y del mismo modo suponga que Y_2 es una variable aleatoria con función de densidad $f_2(y)$ y que $E(Y_2) = \mu_2$ y $\text{Var}(Y_2) = \sigma_2^2$. Suponga que Y es una variable aleatoria cuya densidad es una mezcla de las densidades correspondientes a Y_1 y Y_2. Demuestre que

i $E(Y) = a\mu_1 + (1 - a)\mu_2$.

ii $\text{Var}(Y) = a\sigma_1^2 + (1 - a)\sigma_2^2 + a(1 - a)[\mu_1 - \mu_2]^2$.

[*Sugerencia:* $E(Y_i^2) = \mu_i^2 + \sigma_i^2,\ i = 1, 2.$]

***4.186** Se dice que la variable aleatoria Y, con una función de densidad dada por

$$f(y) = \frac{my^{m-1}}{\alpha}e^{-y^m\alpha}, \qquad 0 \le y < \infty, \alpha,\ m > 0$$

tiene una distribución *Weibull.* La función de densidad Weibull da un buen modelo para la distribución de vida útil para muchos aparatos mecánicos y plantas biológicas y animales. Encuentre la media y la varianza para una variable aleatoria con distribución Weibull con $m = 2$.

***4.187** Consulte el Ejercicio 4.186. Los resistores que se usan en la construcción de un sistema de guía en aviones tienen vidas útiles que siguen una distribución Weibull con $m = 2$ y $\alpha = 10$ (con medidas en miles de horas).

a Encuentre la probabilidad de que la vida útil de un resistor de este tipo seleccionado al azar exceda de 5000 horas.

b Si tres resistores de este tipo operan de manera independiente encuentre la probabilidad de que exactamente uno de los tres se queme antes de 5000 horas de uso.

***4.188** Consulte el Ejercicio 4.186.

a ¿Cuál es el nombre común de la distribución de una variable aleatoria que tiene una distribución Weibull con $m = 1$?

b Obtenga, en términos de los parámetros α y m, la media y la varianza de una variable aleatoria con distribución Weibull.

***4.189** Si $n > 2$ es un entero, la distribución con densidad dada por

$$f(y) = \begin{cases} \dfrac{1}{B(1/2, [n-2]/2)}(1 - y^2)^{(n-4)/2}, & -1 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Se llama la distribución r. Determine la media y la varianza de una variable aleatoria con la distribución r.

***4.190** Una función en ocasiones asociada con variables aleatorias continuas no negativas es la función de porcentaje de falla (o porcentaje de riesgo), que está definida por

$$r(t) = \frac{f(t)}{1 - F(t)}$$

para una función de densidad $f(t)$ con función de distribución correspondiente $F(t)$. Si consideramos la variable aleatoria en cuestión como la duración de un componente, $r(t)$ es proporcional a la probabilidad de falla en un pequeño intervalo después de t, dado que el componente ha seguido en servicio después de t. Demuestre que,

a para una función de densidad exponencial, $r(t)$ es constante,

b para una función de densidad Weibull con $m > 1$, $r(t)$ es una función creciente de t. (Vea el Ejercicio 4.186.)

*__4.191__ Suponga que Y es una variable aleatoria continua con una función de distribución dada por $F(y)$ y función de densidad de probabilidad $f(y)$. Con frecuencia estamos interesados en conocer las probabilidades condicionales de la forma $P(Y \le y | Y \ge c)$ para una constante c.

a Demuestre que, para $y \ge c$,

$$P(Y \le y | Y \ge c) = \frac{F(y) - F(c)}{1 - F(c)}.$$

b Demuestre que la función en el inciso a tiene todas las propiedades de una función de distribución.

c Si la duración Y para una batería tiene una distribución Weibull con $m = 2$ y $\alpha = 3$ (con mediciones en años), encuentre la probabilidad de que la batería dure menos de cuatro años, en vista que ahora ya tiene dos años.

*__4.192__ Las velocidades de partículas de gas pueden ser modeladas por la distribución de Maxwell, cuya función de densidad de probabilidad está dada por

$$f(v) = 4\pi\left(\frac{m}{2\pi KT}\right)^{3/2} v^2 e^{-v^2(m/[2KT])}, \qquad v > 0,$$

donde m es la masa de la partícula, K es la constante de Boltzmann y T es la temperatura absoluta.

a Encuentre la velocidad media de estas partículas.

b La energía cinética de una partícula está dada por $(1/2)mV^2$. Encuentre la energía cinética media para una partícula.

*__4.193__ En vista de que

$$P(Y \le y | Y \ge c) = \frac{F(y) - F(c)}{1 - F(c)}$$

tiene las propiedades de una función de distribución, su derivada tendrá las propiedades de una función de densidad de probabilidad. Esta derivada está dada por

$$\frac{f(y)}{1 - F(c)}, \qquad y \ge c.$$

Podemos entonces hallar el valor esperado de Y, dado que Y es mayor que c, con el uso de

$$E(Y | Y \ge c) = \frac{1}{1 - F(c)} \int_c^\infty y f(y)\, dy.$$

Si Y, la duración de un componente electrónico, tiene una distribución exponencial con media de 100 horas, encuentre el valor esperado de Y, dado que este componente ya ha estado en uso durante 50 horas.

*__4.194__ Podemos demostrar que la función de densidad normal se integra hasta la unidad al demostrar que, si $\mu > 0$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1/2)uy^2}\, dy = \frac{1}{\sqrt{u}}.$$

Esto, a su vez, se puede demostrar al considerar el producto de dos de estas integrales:

$$\frac{1}{2\pi}\left(\int_{-\infty}^{\infty} e^{-(1/2)uy^2}\, dy\right)\left(\int_{-\infty}^{\infty} e^{-(1/2)ux^2}\, dx\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(1/2)u(x^2+y^2)}\, dx\, dy.$$

Al transformar a coordenadas polares, demuestre que la integral doble precedente es igual a $1/u$.

*__4.195__ Sea Z una variable aleatoria normal estándar y $W = (Z^2 + 3Z)^2$.

a Use los momentos de Z (vea Ejercicio 4.199) para obtener la media de W.

b Use el resultado dado en el Ejercicio 4.198 para hallar un valor de w tal que $P(W \le w) \ge .90$.

*__4.196__ Demuestre que $\Gamma(1/2) = \sqrt{\pi}$ al escribir

$$\Gamma(1/2) = \int_0^\infty y^{-1/2} e^{-y}\, dy$$

al hacer la transformación $y = (1/2)x^2$ y emplear el resultado del Ejercicio 4.194.

***4.197** La función $B(\alpha, \beta)$ está definida por

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\, dy.$$

a Si $y = \text{sen}^2\,\theta$, demuestre que

$$B(\alpha, \beta) = 2\int_0^{\pi/2} \text{sen}^{2\alpha-1}\,\theta \cos^{2\beta-1}\,\theta\; d\theta.$$

b Exprese $\Gamma(\alpha)\Gamma(\beta)$ como una integral doble, transforme a coordenadas polares y concluya que

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

***4.198** **La desigualdad Markov** Sea $g(Y)$ una función de la variable aleatoria continua Y, con $E(|g(Y)|) < \infty$. Demuestre que, para toda constante positiva k,

$$P(\,g(Y) \le \;k) \ge 1 - \frac{E(\,g(Y)\,)}{k}.$$

[*Nota:* esta desigualdad también se cumple para variables aleatorias discretas, con una adaptación obvia en la prueba.]

***4.199** Sea Z una variable aleatoria normal estándar.

a Demuestre que los valores esperados de todas las potencias enteras impares de Z son 0. Esto es, si $i = 1, 2, \ldots$, demuestre que $E(Z^{2i-1}) = 0$. [*Sugerencia:* una función $g(\cdot)$ es una *función impar* si, para toda y, $g(-y) = -g(y)$. Para cualquier función impar $g(y)$, $\int_{-\infty}^{\infty} g(y)\, dy = 0$, si la integral existe.]

b Si $i = 1, 2, \ldots$, demuestre que

$$E(Z^{2i}) = \frac{2^i\,\Gamma\left(i + \frac{1}{2}\right)}{\sqrt{\pi}}.$$

[*Sugerencia:* una función $h(\cdot)$ es una *función par* si, para toda y, $h(-y) = h(y)$. Para cualquier función par $h(y)$, $\int_{-\infty}^{\infty} h(y)\, dy = 2\int_0^{\infty} h(y)\, dy$, si las integrales existen. Use este dato, haga el cambio de variable $w = z^2/2$ y use lo que sepa acerca de la función gamma.]

c Use los resultados del inciso b y de los Ejercicios 4.81(b) y 4.194 para deducir $E(Z^2)$, $E(Z^4)$, $E(Z^6)$ y $E(Z^8)$.

d Si $i = 1, 2, \ldots$, demuestre que

$$E(Z^{2i}) = \prod_{j=1}^{i}(2j-1).$$

Esto implica que el i–ésimo momento par es el producto de los primeros i enteros impares.

4.200 Suponga que Y tiene una distribución beta con parámetros α y β.

a Si a es cualquier valor positivo o negativo tal que $\alpha + a > 0$, demuestre que

$$E(Y^a) = \frac{\Gamma(\alpha+a)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+a)}.$$

b ¿Por qué su respuesta al inciso a requirió que $\alpha + a > 0$?

c Demuestre que, con $a = 1$, el resultado en el inciso a da $E(Y) = \alpha/(\alpha + \beta)$.

d Use el resultado del inciso a para dar una expresión para $E(\sqrt{Y})$. ¿Qué necesita para hacer suposiciones acerca de α?

e Use el resultado del inciso a para dar una expresión para $E(1/Y)$, $E(1/\sqrt{Y})$, y $E(1/Y^2)$. ¿Qué necesita para hacer suposiciones acerca de α en cada caso?

CAPÍTULO 5

Distribuciones de probabilidad multivariantes

5.1 Introducción

La intersección de dos o más eventos es frecuentemente de interés para un experimentador. Por ejemplo, un jugador de *black-jack* está interesado en el evento de sacar un as y una “figura” de una baraja de 52 cartas. Un biólogo que observa el número de animales que sobreviven de una camada se preocupa por la intersección de estos eventos:

A: la camada contiene n animales.
B: sobreviven y animales.

Del mismo modo, observar la estatura y peso de una persona representa la intersección de un par específico de eventos asociado con medidas de estatura-peso.

Lo que es más importante para expertos en estadística son las intersecciones que se presentan en el curso de tomar muestras. Suponga que $Y_1, Y_2, \ldots, Y_n$ denota los resultados de n intentos sucesivos de un experimento. Por ejemplo, esta secuencia podría representar los pesos de n personas o las medidas de n características físicas para una sola persona. Un conjunto específico de resultados o mediciones muestrales puede ser expresado en términos de la intersección de los n eventos $(Y_1 = y_1), (Y_2 = y_2), \ldots, (Y_n = y_n)$, que denotaremos como $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ o bien, de un modo más compacto, como $(y_1, y_2, \ldots, y_n)$. El cálculo de la probabilidad de esta intersección es esencial para hacer inferencias acerca de la población de la cual se tomó la muestra y es una razón importante para estudiar distribuciones de probabilidad multivariantes.

5.2 Distribuciones de probabilidad bivariantes y multivariantes

Se pueden definir muchas variables aleatorias sobre el mismo espacio muestral. Por ejemplo, considere el experimento de lanzar un par de dados. El espacio muestral contiene 36 puntos muestrales, correspondientes a las $mn = (6)(6) = 36$ formas en las que pueden aparecer números en las caras de los dados. Cualquiera de las siguientes variables aleatorias podría estar definida sobre el espacio muestral y podría ser de interés para el experimentador:

Y_1: el número de puntos que aparecen en el dado 1.
Y_2: el número de puntos que aparecen en el dado 2.
Y_3: la suma del número de puntos en los dados.
Y_4: el producto del número de puntos que aparecen en los dados.

Los 36 puntos muestrales asociados con el experimento tienen la misma probabilidad y corresponden a los 36 eventos numéricos (y_1, y_2). Así, lanzar un par de números 1 es el evento sencillo (1, 1). Lanzar un 2 en el dado 1 y un 3 en el dado 2 es el evento sencillo (2, 3). Como todos los pares (y_1, y_2) ocurren con la misma frecuencia relativa, asignamos una probabilidad $1/36$ a cada punto muestral. Para este ejemplo sencillo la intersección (y_1, y_2) contiene a lo sumo un punto muestral. En consecuencia, la función de probabilidad bivariante es

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) = 1/36, \qquad y_1 = 1, 2, \ldots, 6, \; y_2 = 1, 2, \ldots, 6.$$

En la Figura 5.1 se muestra una gráfica de la función de probabilidad bivariante para el experimento de lanzar dados. Observe que una probabilidad diferente de cero se asigna a un punto (y_1, y_2) del plano si y sólo si $y_1 = 1, 2, \ldots, 6$ y $y_2 = 1, 2, \ldots, 6$. Entonces, a los 36 puntos del plano se les asignan exactamente probabilidades diferentes de cero. Además, las probabilidades se asignan en tal forma que la suma de las probabilidades diferentes de cero es igual a 1. En la Figura 5.1 los puntos a los que se asignan probabilidades diferentes de cero están representados en el plano (y_1, y_2), mientras que las probabilidades asociadas con estos puntos están dadas por las longitudes de las rectas que aparecen arriba de ellos. La Figura 5.1 puede verse como histograma teórico de frecuencia relativa en tres dimensiones para los pares de observaciones (y_1, y_2). Al igual que en el caso discreto de una sola variable, el histograma teórico da un modelo para el histograma muestral que se obtendría si el experimento de lanzar dados se repitiera un gran número de veces.

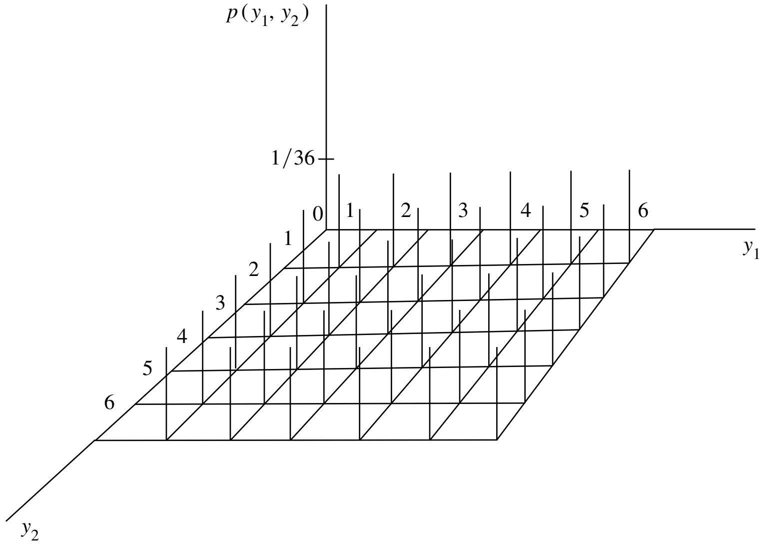


FIGURA **5.1** Función de probabilidad bivariante; y_1 = número de puntos en el dado 1, y_2 = número de puntos en el dado 2

DEFINICIÓN **5.1**

Sean Y_1 y Y_2 variables aleatorias discretas. La *función de probabilidad conjunta* (o bivariante) para Y_1 y Y_2 está dada por

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

En el caso de la variable única que estudiamos en el Capítulo 3 vimos que la función de probabilidad para una variable aleatoria discreta Y asigna probabilidades diferentes de cero a un número finito o contable de valores distintos de Y, en forma tal que la suma de las probabilidades es igual a 1. Del mismo modo, en el caso bivariante la función de probabilidad conjunta $p(y_1, y_2)$ asigna probabilidades diferentes de cero a sólo un número finito o contable de pares de valores (y_1, y_2). Además, las probabilidades diferentes de cero deben sumar 1.

TEOREMA **5.1**

Si Y_1 y Y_2 son variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$, entonces

1. $p(y_1, y_2) \geq 0$ para toda y_1, y_2.
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$, donde la suma es para todos los valores (y_1, y_2) a los que se asignan probabilidades diferentes de cero.

Al igual que en el caso discreto univariante, la función de probabilidad conjunta para variables aleatorias discretas a veces se denomina *función de masa de probabilidad conjunta* porque especifica la probabilidad (masa) asociada con cada uno de los posibles pares de valores para las variables aleatorias. Una vez que la función de probabilidad conjunta se haya determinado para variables aleatorias discretas Y_1 y Y_2, calcular las probabilidades conjuntas en donde aparecen Y_1 y Y_2 es fácil. Para el experimento de lanzar dados, $P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2)$ es

$$\begin{aligned} P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2) &= p(2, 1) + p(2, 2) + p(3, 1) + p(3, 2) \\ &= 4/36 = 1/9. \end{aligned}$$

EJEMPLO **5.1** Un supermercado local tiene tres cajas. Dos clientes llegan a las cajas en momentos diferentes cuando las cajas no atienden a otros clientes. Cada cliente escoge una caja de manera aleatoria, independientemente del otro. Denote con Y_1 el número de clientes que escogen la caja 1 y con Y_2 el número que selecciona la caja 2. Encuentre la función de probabilidad conjunta de Y_1 y Y_2.

Solución Podríamos proceder en muchas formas. La más directa es considerar el espacio muestral asociado con el experimento. Denotemos con el par $\{i, j\}$ el evento sencillo de que el primer cliente escogió la caja i y el segundo cliente escogió la caja j, donde $i, j = 1, 2$ y 3. Usando la regla *mn*, el espacio muestral está formado por $3 \times 3 = 9$ puntos muestrales. De acuerdo con las suposiciones dadas antes, cada punto muestral es igualmente probable y tiene probabilidad $1/9$. El espacio muestral asociado con el experimento es

$$S = [\{1, 1\}, \{1, 2\}, \{1, 3\}, \{2, 1\}, \{2, 2\}, \{2, 3\}, \{3, 1\}, \{3, 2\}, \{3, 3\}].$$

Observe que el punto muestral $\{1, 1\}$ es el único correspondiente a $(Y_1 = 2, Y_2 = 0)$ y por tanto $P(Y_1 = 2, Y_2 = 0) = 1/9$. Del mismo modo, $P(Y_1 = 1, Y_2 = 1) = P(\{1, 2\} \text{ o } \{2, 1\}) = 2/9$. La Tabla 5.1 contiene las probabilidades asociadas con cada posible par de valores para Y_1 y Y_2, es decir, la función de probabilidad conjunta para Y_1 y Y_2. Como siempre, los resultados del Teorema 5.1 se cumplen para este ejemplo.

Tabla 5.1 Función de probabilidad para Y_1 y Y_2, Ejemplo 5.1

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

■

Al igual que en el caso de variables aleatorias univariantes, la distinción entre variables aleatorias continuas conjuntas y discretas conjuntas puede ser caracterizado en términos de sus funciones de distribución (conjuntas).

DEFINICIÓN **5.2** Para cualesquiera variables aleatorias Y_1 y Y_2, la función de distribución (bivariante) conjunta $F(y_1, y_2)$ es

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

Para dos variables discretas Y_1 y Y_2, $F(y_1, y_2)$ está dada por

$$F(y_1, y_2) = \sum_{t_1 \le y_1} \sum_{t_2 \le y_2} p(t_1, t_2).$$

Para el experimento de lanzar un dado,

$$\begin{aligned} F(2, 3) &= P(Y_1 \le 2, Y_2 \le 3) \\ &= p(1, 1) + p(1, 2) + p(1, 3) + p(2, 1) + p(2, 2) + p(2, 3). \end{aligned}$$

Como $p(y_1, y_2) = 1/36$ para todos los pares de valores de y_1 y y_2 en consideración, $F(2, 3) = 6/36 = 1/6$.

EJEMPLO **5.2** Considere las variables aleatorias Y_1 y Y_2 del Ejemplo 5.1. Encuentre $F(-1, 2)$, $F(1.5, 2)$ y $F(5, 7)$.

Solución Usando los resultados de la Tabla 5.1 vemos que

$$F(-1, 2) = P(Y_1 \le -1, Y_2 \le 2) = P(\emptyset) = 0.$$

Además,

$$\begin{aligned} F(1.5, 2) &= P(Y_1 \le 1.5, Y_2 \le 2) \\ &= p(0, 0) + p(0, 1) + p(0, 2) + p(1, 0) + p(1, 1) + p(1, 2) = 8/9. \end{aligned}$$

De manera similar,

$$F(5, 7) = P(Y_1 \le 5, Y_2 \le 7) = 1.$$

Observe que $F(y_1, y_2) = 1$ para toda y_1, y_2 tal que $\text{mín}\{y_1, y_2\} \ge 2$. También, $F(y_1, y_2) = 0$ si $\text{mín}\{y_1, y_2\} < 0$. ■

Se dice que dos variables aleatorias son continuas conjuntas si su función de distribución conjunta $F(y_1, y_2)$ es continua en ambos argumentos.

DEFINICIÓN **5.3**

Sean Y_1 y Y_2 variables aleatorias continuas con función de distribución conjunta $F(y_1, y_2)$. Si existe una función no negativa $f(y_1, y_2)$, tal que

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_2\, dt_1,$$

para toda $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$, entonces se dice que Y_1 y Y_2 son *variables aleatorias continuas conjuntas*. La función $f(y_1, y_2)$ recibe el nombre de *función de densidad de probabilidad conjunta.*

Las funciones de distribución acumulativa bivariante satisfacen un conjunto de propiedades similares a las especificadas para funciones de distribución acumulativa univariante.

TEOREMA 5.2

Si Y_1 y Y_2 son variables aleatorias con función de distribución conjunta $F(y_1, y_2)$, entonces

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. Si $y_1^* \geq y_1$ y $y_2^* \geq y_2$, entonces

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0.$$

La parte 3 resulta de que

$$\begin{aligned} &F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \\ &\quad = P(y_1 < Y_1 \leq y_1^*, y_2 < Y_2 \leq y_2^*) \geq 0. \end{aligned}$$

Observe que $F(\infty, \infty) \equiv \text{lím}_{y_1 \to \infty} \text{lím}_{y_2 \to \infty} F(y_1, y_2) = 1$ implica que la función de densidad conjunta $f(y_1, y_2)$ debe ser tal que la integral de $f(y_1, y_2)$ para todos los valores de (y_1, y_2) es 1.

TEOREMA 5.3

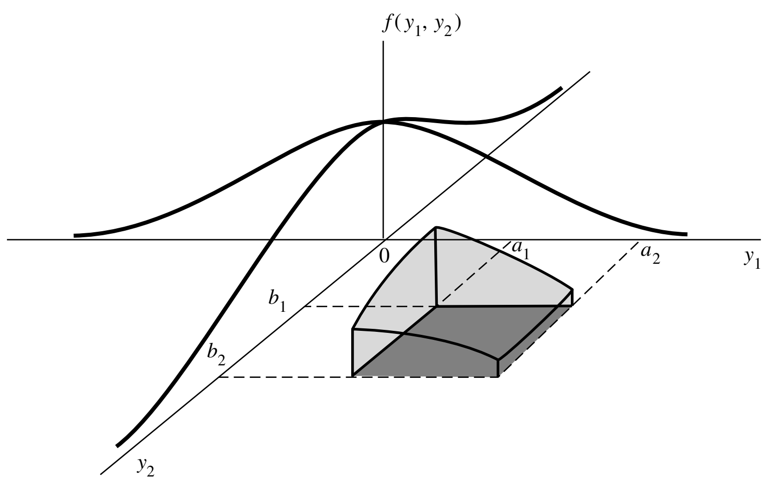
Si Y_1 y Y_2 son variables aleatorias continuas conjuntas con una función de densidad conjunta dada por $f(y_1, y_2)$, entonces

1. $f(y_1, y_2) \geq 0$ para toda y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = 1$.

Al igual que en el caso continuo univariante que se estudia en el Capítulo 4, la función de densidad conjunta puede ser interpretada de manera intuitiva como un modelo para el histograma de frecuencia relativa conjunta para Y_1 y Y_2.

Para el caso continuo univariante, las áreas bajo la densidad de probabilidad para un intervalo corresponden a probabilidades. De igual manera, la función de densidad de probabilidad bivariante $f(y_1, y_2)$ traza una superficie de densidad de probabilidad sobre el plano (y_1, y_2) (Figura 5.2).

FIGURA **5.2** Función de densidad bivariante $f(y_1, y_2)$



Los volúmenes bajo esta superficie representan probabilidades. Así, $P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2)$ es el volumen sombreado que se ve en la Figura 5.2 y es igual a

$$\int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2)\, dy_1\, dy_2.$$

EJEMPLO **5.3** Suponga que una partícula radiactiva se localiza aleatoriamente en un cuadrado con lados de longitud unitaria. Esto es, si se consideran dos regiones de igual área y dentro del cuadrado unitario es igualmente probable que la partícula se encuentre en cualquiera de las dos. Denote con Y_1 y Y_2 las coordenadas de la ubicación de la partícula. Un modelo razonable para el histograma de frecuencia relativa para Y_1 y Y_2 es la análoga bivariante de la función de densidad uniforme univariante:

$$f(y_1, y_2) = \begin{cases} 1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Trace la superficie de densidad de probabilidad.
b Encuentre $F(.2, .4)$.
c Encuentre $P(.1 \leq Y_1 \leq .3, 0 \leq Y_2 \leq .5)$.

Solución **a** El trazo se muestra en la Figura 5.3.

b

$$\begin{aligned} F(.2, .4) &= \int_{-\infty}^{.4} \int_{-\infty}^{.2} f(y_1, y_2)\, dy_1\, dy_2 \\ &= \int_{0}^{.4} \int_{0}^{.2} (1)\, dy_1\, dy_2 \\ &= \int_{0}^{.4} \left(y_1 \Big]_0^{.2} \right) dy_2 = \int_0^{.4} .2\, dy_2 = .08. \end{aligned}$$

La probabilidad $F(.2, .4)$ corresponde al volumen bajo $f(y_1, y_2) = 1$, que está sombreado en la Figura 5.3. Como lo indican consideraciones geométricas, la probabilidad deseada (volumen) es igual a .08, que obtuvimos mediante integración al principio de esta sección.

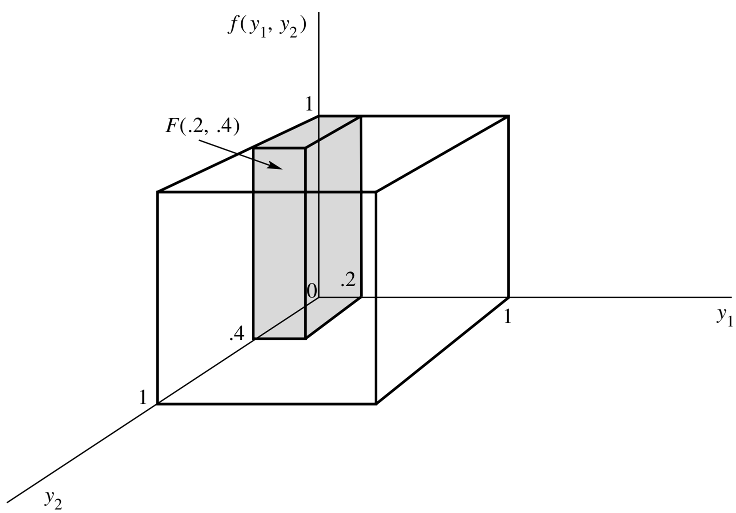


FIGURA **5.3** Representación geométrica de $f(y_1, y_2)$, Ejemplo 5.3

c
$$P(.1 \leq Y_1 \leq .3, 0 \leq Y_2 \leq .5) = \int_0^{.5} \int_{.1}^{.3} f(y_1, y_2)\, dy_1\, dy_2$$
$$= \int_0^{.5} \int_{.1}^{.3} 1\, dy_1\, dy_2 = .10.$$

Esta probabilidad corresponde al volumen bajo la función de densidad $f(y_1, y_2) = 1$ que está arriba de la región $.1 \leq y_1 \leq .3$, $0 \leq y_2 \leq .5$. Al igual que la solución del inciso b, la solución actual se puede obtener con el uso de conceptos de geometría elemental. La densidad o altura de la superficie es igual a 1 y por tanto la probabilidad deseada (volumen) es

$$P(.1 \leq Y_1 \leq .3, 0 \leq Y_2 \leq .5) = (.2)(.5)(1) = .10.$$ ■

En el siguiente ejemplo se ilustra un modelo bivariante ligeramente más complicado.

EJEMPLO **5.4** Se ha de almacenar gasolina en un enorme tanque una vez al principio de cada semana y luego se vende a clientes individuales. Denote con Y_1 el nivel de gasolina (proporción) que alcanza el tanque después de surtirlo. Debido a suministros limitados, Y_1 varía de una semana a otra. Denote con Y_2 la proporción de la capacidad del tanque que se vende durante la semana. Como Y_1 y Y_2 son proporciones, estas dos variables toman valores entre 0 y 1. Además, la cantidad de gasolina vendida, y_2, no puede ser mayor que la cantidad disponible, y_1. Suponga que la función de densidad conjunta para Y_1 y Y_2 está dada por

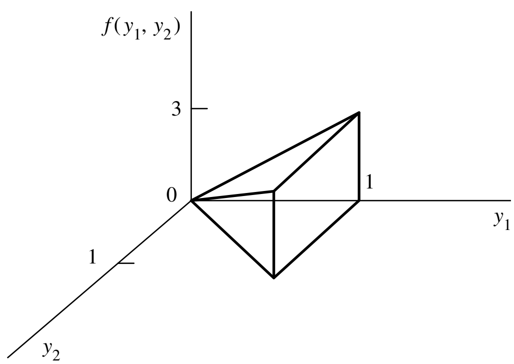
$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \leq y_2 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En la Figura 5.4 se muestra una gráfica de esta función.

Encuentre la probabilidad de que menos de la mitad del tanque tenga gasolina y más de un cuarto del tanque se venda.

Solución Buscamos $P(0 \leq Y_1 \leq .5, Y_2 > .25)$. Para cualquier variable aleatoria continua, la probabilidad de observar un valor en una región es el volumen bajo la función de densidad por arriba de la región de interés. La función de densidad $f(y_1, y_2)$ es positiva sólo en la región triangular

FIGURA **5.4**
Función de densidad conjunta para el Ejemplo 5.4



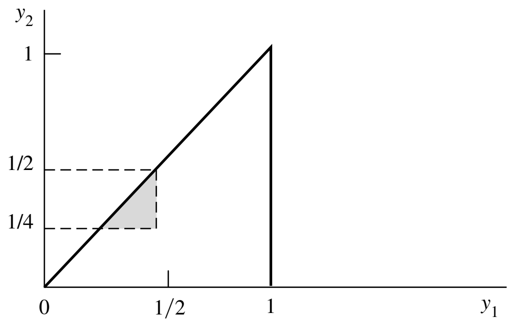


FIGURA **5.5** Región de integración para el Ejemplo 5.4

grande del plano (y_1, y_2) que se ve en la Figura 5.5. Estamos interesados sólo en valores de y_1 y y_2 tales que $0 \le y_1 \le .5$ y $y_2 > .25$. La intersección de esta región y la región donde la función de densidad es positiva está dada por el pequeño triángulo (sombreado) de la Figura 5.5. En consecuencia, la probabilidad que deseamos es el volumen bajo la función de densidad de la Figura 5.4 arriba de la región sombreada del plano (y_1, y_2) que se ve en la Figura 5.5.

Entonces, tenemos

$$
\begin{aligned}
P(0 \le Y_1 \le .5,\ \ .25 \le Y_2) &= \int_{1/4}^{1/2} \int_{1/4}^{y_1} 3y_1 \, dy_2 \, dy_1 \\
&= \int_{1/4}^{1/2} 3y_1 \left(y_2 \Big]_{1/4}^{y_1} \right) dy_1 \\
&= \int_{1/4}^{1/2} 3y_1(y_1 - 1/4) \, dy_1 \\
&= \left[y_1^3 - (3/8)y_1^2 \right] \Big]_{1/4}^{1/2} \\
&= [(1/8) - (3/8)(1/4)] - [(1/64) - (3/8)(1/16)] \\
&= 5/128.
\end{aligned}
$$

■

El cálculo de la probabilidad especificada en el Ejemplo 5.4 comprendió integrar la función de densidad conjunta para Y_1 y Y_2 sobre la región apropiada. La especificación de los límites de integración se hizo más fácil al trazar la región de integración en la Figura 5.5. Este método, trazando la región apropiada de integración, con frecuencia facilita establecer la integral apropiada.

Los métodos estudiados en esta sección se pueden usar para calcular la probabilidad de la intersección de dos eventos ($Y_1 = y_1$, $Y_2 = y_2$). De igual modo podemos definir una función de probabilidad (o función de densidad de probabilidad) para la intersección de n eventos ($Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$). La función de probabilidad conjunta correspondiente al caso discreto está dada por

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n).$$

La función de densidad conjunta de $Y_1, Y_2, \ldots, Y_n$ está dada por $f(y_1, y_2, \ldots, y_n)$. Al igual que en el caso bivariante, estas funciones dan modelos para las distribuciones de frecuencia relativa conjunta de las poblaciones de observaciones conjuntas ($y_1, y_2, \ldots, y_n$) para el caso discreto y el caso continuo, respectivamente. En el caso continuo,

$$P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) = F(y_1, \ldots, y_n)$$
$$= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) dt_n \ldots dt_1$$

para todo conjunto de números reales ($y_1, y_2, \ldots, y_n$). Las funciones de distribución multivariantes definidas por esta igualdad satisfacen propiedades semejantes a las especificadas para el caso bivariante.

Ejercicios

5.1 Los contratos para dos trabajos de construcción se asignan aleatoriamente a una o más de tres empresas, A, B y C. Denote con Y_1 el número de contratos asignados a la empresa A y Y_2 el número de contratos asignados a la B. Recuerde que cada empresa puede recibir 0, 1 o 2 contratos.

a Encuentre la función de probabilidad conjunta para Y_1 y Y_2.

b Encuentre $F(1, 0)$.

5.2 Tres monedas balanceadas se lanzan en forma independiente al aire. Una de las variables de interés es Y_1, el número de caras. Denote con Y_2 la cantidad de dinero ganado en una apuesta colateral en la siguiente forma. Si la primera cara aparece en el primer tiro, usted gana \$1. Si la primera cara aparece en el tiro 2 o en el 3 usted gana \$2 o \$3, respectivamente. Si no aparece una cara, usted pierde \$1 (esto es, gana –\$1).

a Encuentre la función de probabilidad conjunta para Y_1 y Y_2.

b ¿Cuál es la probabilidad de que haya menos de tres caras y usted gane \$1 o menos? [Esto es, encuentre $F(2, 1)$.]

5.3 De nueve ejecutivos de una empresa financiera, cuatro están casados, tres nunca se han casado y dos están divorciados. Tres de los ejecutivos se han de seleccionar para un ascenso. Denote con Y_1 el número de ejecutivos casados y con Y_2 el número de ejecutivos que no se han casado entre los tres seleccionados para el cargo. Suponiendo que los tres se seleccionan aleatoriamente de entre los nueve disponibles, encuentre la función de probabilidad conjunta de Y_1 y Y_2.

5.4 A continuación se da la función de probabilidad conjunta asociada con datos obtenidos en un estudio de accidentes automovilísticos en los que un niño (de menos de 5 años de edad) estaba en el auto y hubo al menos una persona muerta. Específicamente, el estudio se concentró en si el niño sobrevivió y qué tipo de cinturón de seguridad (si lo había) utilizaba. Defina

$$Y_1 = \begin{cases} 0, & \text{si el niño sobrevivió,} \\ 1, & \text{si no,} \end{cases} \quad \text{y} \quad Y_2 = \begin{cases} 0, & \text{si no usaba cinturón,} \\ 1, & \text{si usaba cinturón para adulto,} \\ 2, & \text{si usaba cinturón del asiento del auto.} \end{cases}$$

Observe que Y_1 es el número de fallecimientos por niño y, como los asientos para niños del auto tienen por lo general dos cinturones, Y_2 es el número de cinturones de seguridad que se usaban en el momento del accidente.

	y_1		
y_2	0	1	Total
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
Total	.76	.24	1.00

a Verifique que la función de probabilidad precedente satisface al Teorema 5.1.

b Encuentre $F(1, 2)$. ¿Cuál es la interpretación de este valor?

5.5 Consulte el Ejemplo 5.4. La densidad conjunta de Y_1, la proporción de la capacidad del tanque que se abastece al principio de la semana, y Y_2, la proporción de la capacidad vendida durante la semana, está dada por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $F(1/2, 1/3) = P(Y_1 \le 1/2, Y_2 \le 1/3)$.

b Encuentre $P(Y_2 \le Y_1/2)$, la probabilidad de que la cantidad vendida sea menor que la mitad de la cantidad comprada.

5.6 Consulte el Ejemplo 5.3. Si una partícula radiactiva se selecciona aleatoriamente en un cuadrado de longitud unitaria, un modelo razonable para la función de densidad conjunta para Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Cuál es $P(Y_1 - Y_2 > .5)$?

b ¿Cuál es $P(Y_1Y_2 < .5)$?

5.7 Supongamos que Y_1 y Y_2 tienen la función de densidad conjunta

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Cuál es $P(Y_1 < 1, y_2 > 5)$?

b ¿Cuál es $P(Y_1 + Y_2 < 3)$?

5.8 Tengan Y_1 y Y_2 una función de densidad de probabilidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} ky_1y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de k que haga de ésta una función de densidad de probabilidad.

b Encuentre la función de distribución conjunta para Y_1 y Y_2.

c Encuentre $P(Y_1 \le 1/2, Y_2 \le 3/4)$.

5.9 Tengan Y_1 y Y_2 una función de densidad de probabilidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} k(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el valor de k que haga de ésta una función de densidad de probabilidad.

b Encuentre $P(Y_1 \le 3/4, Y_2 \ge 1/2)$.

5.10 Un ingeniero ambiental mide la cantidad (en peso) de partículas contaminantes en muestras de aire de cierto volumen recolectado en dos chimeneas en una planta de energía alimentada con carbón. Una de las chimeneas está equipada con un aparato limpiador. Denote con Y_1 la cantidad de contaminante por muestra recolectada arriba de la chimenea que no tiene aparato limpiador y denote con Y_2 la cantidad de contaminante por muestra recolectada arriba de la chimenea que está equipada con el aparato limpiador.

Suponga que el comportamiento de frecuencia relativa de Y_1 y Y_2 puede ser modelado por

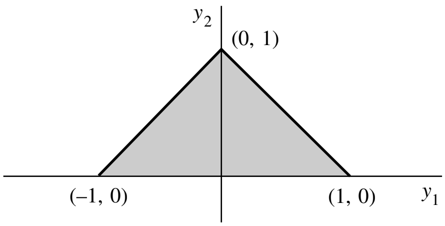
$$f(y_1, y_2) = \begin{cases} k, & 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 1, 2y_2 \leq y_1 \\ 0, & \textit{en cualquier otro punto}. \end{cases}$$

Esto es, Y_1 y Y_2 están uniformemente distribuidas sobre la región dentro del triángulo limitado por $y_1 = 2$, $y_2 = 0$ y $2y_2 = y_1$.

a Encuentre el valor de k que haga de ésta una función de densidad de probabilidad.

b Encuentre $P(Y_1 \geq 3Y_2)$. Esto es, encuentre la probabilidad de que el aparato limpiador reduzca la cantidad de contaminante en un tercio o más.

5.11 Suponga que Y_1 y Y_2 están uniformemente distribuidas sobre el triángulo sombreado del siguiente diagrama.



a Encuentre $P(Y_1 \leq 3/4, Y_2 \leq 3/4)$.

b Encuentre $P(Y_1 - Y_2 \geq 0)$.

5.12 Denote con Y_1 y Y_2 las proporciones de dos tipos diferentes de componentes en una muestra proveniente de una mezcla de productos químicos usados como insecticida. Suponga que Y_1 y Y_2 tienen la función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 2, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, 0 \leq y_1 + y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(Observe que $Y_1 + Y_2 \leq 1$ porque las variables aleatorias denotan proporciones dentro de la misma muestra.) Encuentre

a $P(Y_1 \leq 3/4, Y_2 \leq 3/4)$.

b $P(Y_1 \leq 1/2, Y_2 \leq 1/2)$.

5.13 La función de densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \leq y_2 \leq 1 - y_1, 0 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $F(1/2, 1/2)$.

b Encuentre $F(1/2, 2)$.

c Encuentre $P(Y_1 > Y_2)$.

5.14 Suponga que las variables aleatorias Y_1 y Y_2 tienen función de densidad de probabilidad conjunta $f(y_1, y_2)$ dada por

$$f(y_1, y_2) = \begin{cases} 6y_1^2y_2, & 0 \leq y_1 \leq y_2, y_1 + y_2 \leq 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Verifique que ésta es una función de densidad conjunta válida.

b ¿Cuál es la probabilidad de que $Y_1 + Y_2$ sea menor que 1?

5.15 La administración en un restaurante de comida rápida está interesada en el comportamiento conjunto de las variables aleatorias Y_1, definidas como el tiempo total entre la llegada de un cliente a la tienda y la salida de la ventanilla de servicio y Y_2, el tiempo que un cliente espera en la fila antes de llegar a la ventanilla de servicio. Como Y_1 incluye el tiempo que un cliente espera en la fila, debemos tener $Y_1 \geq Y_2$. La distribución de frecuencia relativa de valores observados de Y_1 y Y_2 puede ser modelada por la función de densidad de probabilidad

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \leq y_2 \leq y_1 < \infty, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

con el tiempo medido en minutos. Encuentre.

a $P(Y_1 < 2, Y_2 > 1)$.

b $P(Y_1 \geq 2Y_2)$.

c $P(Y_1 - Y_2 \geq 1)$. (Observe que $Y_1 - Y_2$ denota el tiempo que se pasa en la ventanilla de servicio.)

5.16 Denote con Y_1 y Y_2 las proporciones de tiempo (en un día hábil) durante las cuales los empleados I y II, respectivamente, realizan sus tareas asignadas. El comportamiento de frecuencia relativa conjunta de Y_1 y Y_2 está modelado por la función de densidad

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $P(Y_1 < 1/2, Y_2 > 1/4)$.

b Encuentre $P(Y_1 + Y_2 \leq 1)$.

5.17 Denote con (Y_1, Y_2) las coordenadas de un punto seleccionado aleatoriamente dentro de un círculo unitario cuyo centro está en el origen. Esto es, Y_1 y Y_2 tienen una función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} \dfrac{1}{\pi}, & y_1^2 + y_2^2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $P(Y_1 \leq Y_2)$.

5.18 Un sistema electrónico tiene uno de cada dos tipos diferentes de componentes en operación conjunta. Denote con Y_1 y Y_2 las duraciones aleatorias de los componentes del tipo I y tipo II, respectivamente. La función de densidad conjunta está dada por

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(Las mediciones son en cientos de horas.) Encuentre $P(Y_1 > 1, Y_2 > 1)$.

5.3 Distribuciones de probabilidad marginal y condicional

Recuerde que los valores distintos tomados por una variable aleatoria discreta representan eventos mutuamente excluyentes. De manera análoga, para todos los distintos pares de valores y_1, y_2, los eventos bivariantes $(Y_1 = y_1, Y_2 = y_2)$, representados por (y_1, y_2), son eventos mutuamente excluyentes. Se deduce que el evento univariante $(Y_1 = y_1)$ es la unión de eventos bivariantes del tipo $(Y_1 = y_1, Y_2 = y_2)$, con la unión tomada para todos los posibles valores de y_2.

Por ejemplo, reconsidere el experimento de tirar un dado de la Sección 5.2, donde

$$Y_1 = \text{número de puntos de la cara superior del dado 1,}$$
$$Y_2 = \text{número de puntos de la cara superior del dado 2.}$$

Entonces

$$\begin{aligned}
P(Y_1 = 1) &= p(1, 1) + p(1, 2) + p(1, 3) + \ldots + p(1, 6) \\
&= 1/36 + 1/36 + 1/36 + \ldots + 1/36 = 6/36 = 1/6 \\
P(Y_1 = 2) &= p(2, 1) + p(2, 2) + p(2, 3) + \ldots + p(2, 6) = 1/6 \\
&\vdots \\
P(Y_1 = 6) &= p(6, 1) + p(6, 2) + p(6, 3) + \ldots + p(6, 6) = 1/6.
\end{aligned}$$

Expresadas en notación de sumatoria, las probabilidades acerca de la variable Y_1 sola son

$$P(Y_1 = y_1) = p_1(y_1) = \sum_{y_2=1}^{6} p(y_1, y_2).$$

Del mismo modo, las probabilidades correspondientes a valores de la variable Y_2 sola están dadas por

$$p_2(y_2) = P(Y_2 = y_2) = \sum_{y_1=1}^{6} p(y_1, y_2).$$

La sumatoria en el caso discreto corresponde a la integración en el caso continuo, que nos lleva a la siguiente definición.

DEFINICIÓN **5.4**

a Sean Y_1 y Y_2 variables aleatorias discretas conjuntas con función de probabilidad $p(y_1, y_2)$. Entonces las *funciones de probabilidad marginal* de Y_1 y Y_2, respectivamente, están dadas por

$$p_1(y_1) = \sum_{\text{todos } y_2} p(y_1, y_2) \quad \text{y} \quad p_2(y_2) = \sum_{\text{todos } y_1} p(y_1, y_2).$$

b Sean Y_1 y Y_2 variables aleatorias continuas conjuntas con función de densidad conjunta $f(y_1, y_2)$. Entonces las *funciones de densidad marginal* de Y_1 y Y_2, respectivamente, están dadas por

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 \quad \text{y} \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1.$$

El término *marginal*, como se aplica a las funciones de probabilidad univariante de Y_1 y Y_2, tiene significado intuitivo. Para hallar $p_1(y_1)$, sumamos $p(y_1, y_2)$ para todos los valores de y_2 y por tanto acumulamos las probabilidades en el eje y_1 (o margen). Los casos discretos y continuos se ilustran en los siguientes dos ejemplos.

EJEMPLO **5.5** De un grupo de tres republicanos, dos demócratas y uno independiente se ha de seleccionar aleatoriamente un comité de dos personas. Denote con Y_1 el número de republicanos y con Y_2 el número de demócratas del comité. Encuentre la función de probabilidad conjunta de Y_1 y Y_2 y luego encuentre la función de probabilidad marginal de Y_1.

Solución Las probabilidades buscadas aquí son semejantes a las probabilidades hipergeométricas del Capítulo 3. Por ejemplo,

$$P(Y_1 = 1, Y_2 = 1) = p(1,\ 1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{0}}{\binom{6}{2}} = \frac{3(2)}{15} = \frac{6}{15}$$

debido a que hay 15 puntos muestrales igualmente probables; para el evento en cuestión debemos seleccionar un republicano de entre los tres, un demócrata de entre los dos y cero independientes. Cálculos semejantes llevan a las otras probabilidades que se ven en la Tabla 5.2.

Para hallar $p_1(y_1)$, debemos sumar los valores de Y_2, como indica la Definición 5.4. Por tanto, estas probabilidades están dadas por los totales de columna de la Tabla 5.2. Esto es,

$$p_1(0) = p(0,\ 0) + p(0,\ 1) + p(0,\ 2) = 0 + 2/15 + 1/15 = 3/15.$$

Del mismo modo,

$$p_1(1) = 9/15 \quad \text{y} \quad p_1(2) = 3/15.$$

En forma análoga, la función de probabilidad marginal de Y_2 está dada por los totales de fila.

Tabla 5.2 Función de probabilidad conjunta para Y_1 y Y_2, Ejemplo 5.5

	y_1			
y_2	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

■

EJEMPLO **5.6** Sea

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Grafique $f(y_1,\ y_2)$ y encuentre las funciones de densidad marginal para Y_1 y Y_2.

Solución Geométricamente, $f(y_1, y_2)$ describe una superficie en forma de cuña, como se ve en la Figura 5.6.

Antes de aplicar la definición 5.4 para hallar $f_1(y_1)$ y $f_2(y_2)$, usaremos la Figura 5.6 para visualizar el resultado. Si la probabilidad representada por la cuña estuviera acumulada en el eje y_1 (acumulando probabilidad a lo largo de líneas paralelas al eje y_2), el resultado sería una

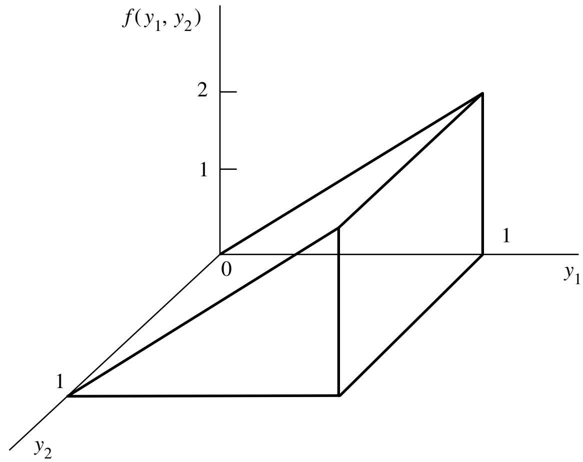


FIGURA **5.6** Representación geométrica de $f(y_1, y_2)$, Ejemplo 5.6

densidad de probabilidad triangular que se vería como el lado de la cuña de la Figura 5.6. Si la probabilidad estuviera acumulada a lo largo del eje y_2 (acumulándose a lo largo de líneas paralelas al eje y_1), la densidad resultante sería uniforme. Confirmaremos estas soluciones visuales mediante la aplicación de la Definición 5.4. Entonces, si $0 \leq y_1 \leq 1$,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 = \int_0^1 2y_1\, dy_2 = 2y_1 \left(y_2 \Big]_0^1 \right)$$

y si $y_1 < 0$ o $y_1 > 1$,

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 = \int_0^1 0\, dy_2 = 0.$$

Entonces,

$$f_1(y_1) = \begin{cases} 2y_1, & 0 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Del mismo modo, si $0 \leq y_2 \leq 1$,

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 = \int_0^1 2y_1\, dy_1 = y_1^2 \Big]_0^1 = 1$$

y si $y_2 < 0$ o $y_2 > 1$,

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1 = \int_0^1 0\, dy_1 = 0.$$

Resumiendo,

$$f_2(y_2) = \begin{cases} 1, & 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Las gráficas de $f_1(y_1)$ y $f_2(y_2)$ trazan densidades de probabilidad triangulares y uniformes, respectivamente, como es de esperarse. ■

Llevemos ahora nuestra atención a distribuciones condicionales, viendo primero al caso discreto.

La ley multiplicativa (Sección 2.8) da la probabilidad de la intersección $A \cap B$ como

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

donde $P(A)$ es la probabilidad incondicional de A y $P(B|A)$ es la probabilidad de B dado que A ha ocurrido. Ahora considere la intersección de los dos eventos numéricos, $(Y_1 = y_1)$ y $(Y_2 = y_2)$, representada por el evento bivariante (y_1, y_2). Se deduce directamente de la ley multiplicativa de probabilidad que la probabilidad bivariante para la intersección (y_1, y_2) es

$$\begin{aligned} p(y_1, y_2) &= p_1(y_1)p(y_2|y_1) \\ &= p_2(y_2)p(y_1|y_2). \end{aligned}$$

Las probabilidades $p_1(y_1)$ y $p_2(y_2)$ están asociadas con las distribuciones de probabilidad univariantes para Y_1 y Y_2 individualmente (recuerde el Capítulo 3). Usando la interpretación de probabilidad condicional estudiada en el Capítulo 2, $p(y_1|y_2)$ es la probabilidad de que la variable aleatoria Y_1 sea igual a y_1, dado que Y_2 toma el valor y_2.

DEFINICIÓN **5.5**

Si Y_1 y Y_2 son variables aleatorias discretas conjuntas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$, respectivamente, entonces la *función de probabilidad discreta condicional* de Y_1 dada Y_2 es

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)},$$

siempre que $p_2(y_2) > 0$.

Entonces, $P(Y_1 = 2|Y_2 = 3)$ es la probabilidad condicional de que $Y_1 = 2$ dado que $Y_2 = 3$. Una interpretación similar se puede unir a la probabilidad condicional $p(y_2|y_1)$. Observe que $p(y_1|y_2)$ es *indefinida* si $p_2(y_2) = 0$.

EJEMPLO **5.7** Consulte el Ejemplo 5.5 y encuentre la distribución condicional de Y_1 dado que $Y_2 = 1$. Esto es, dado que una de las dos personas del comité es demócrata, encuentre la distribución condicional para el número de republicanos seleccionados para el comité.

Solución Las probabilidades conjuntas están dadas en la Tabla 5.2. Para hallar $p(y_1|Y_2 = 1)$, nos concentramos en la fila correspondiente a $Y_2 = 1$. Entonces

$$P(Y_1 = 0|Y_2 = 1) = \frac{p(0, 1)}{p_2(1)} = \frac{2/15}{8/15} = \frac{1}{4},$$

$$P(Y_1 = 1|Y_2 = 1) = \frac{p(1, 1)}{p_2(1)} = \frac{6/15}{8/15} = \frac{3}{4},$$

y

$$P(Y_1 \geq 2|Y_2 = 1) = \frac{p(2, 1)}{p_2(1)} = \frac{0}{8/15} = 0.$$

En el comité seleccionado aleatoriamente, si una persona es demócrata (o, lo que es lo mismo, si $Y_2 = 1$), hay una alta probabilidad de que el otro sea republicano (o sea $Y_1 = 1$).

■

En el caso continuo podemos obtener una analogía apropiada de la función de probabilidad condicional $p(y_1|y_2)$, pero no se obtiene en una forma tan sencilla. Si Y_1 y Y_2 son continuas, $P(Y_1 = y_1|Y_2 = y_2)$ no se puede definir como en el caso discreto porque $(Y_1 = y_1)$ y $(Y_2 = y_2)$ son eventos con probabilidad cero. Las siguientes consideraciones, sin embargo, llevan a una definición útil y consistente para una función de densidad condicional.

Suponiendo que Y_1 y Y_2 son continuas conjuntas con función de densidad $f(y_1, y_2)$, podríamos estar interesados en una probabilidad de la forma

$$P(Y_1 \leq y_1|Y_2 = y_2) = F(y_1|y_2),$$

que, como función de y_1 para una y_2 fija, se denomina *función de distribución condicional* de Y_1, dado que $Y_1 = y_2$.

DEFINICIÓN **5.6**

Si Y_1 y Y_2 son variables aleatorias continuas conjuntas con función de densidad conjunta $f(y_1, y_2)$, entonces la *función de distribución condicional* de Y_1 dado que $Y_2 = y_2$ es

$$F(y_1|y_2) = P(Y_1 \leq y_1|Y_2 = y_2).$$

Observe que $F(y_1|y_2)$ es una función de y_1 para un valor fijo de y_2.

Si pudiéramos tomar $F(y_1|y_2)$, multiplicarlo por $P(Y_2 = y_2)$ para cada posible valor de Y_2 y sumar todas las probabilidades resultantes, podríamos obtener $F(y_1)$. Esto no es posible porque el número de valores para y_2 es incontable y todas las probabilidades $P(Y_2 = y_2)$ son cero. Pero podemos hacer algo análogo al multiplicarlo por $f_2(y_2)$ y luego integrar para obtener

$$F(y_1) = \int_{-\infty}^{\infty} F(y_1|y_2) f_2(y_2)\, dy_2.$$

La cantidad $f_2(y_2)dy_2$ se puede considerar como la probabilidad aproximada de que Y_2 tome un valor en un pequeño intervalo alrededor de y_2, y la integral es una suma generalizada.

Ahora, de consideraciones previas, sabemos que

$$\begin{aligned} F(y_1) &= \int_{-\infty}^{y_1} f_1(t_1)\, dt_1 = \int_{-\infty}^{y_1} \left[\int_{-\infty}^{\infty} f(t_1, y_2)\, dy_2\right] dt_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1\, dy_2. \end{aligned}$$

De estas dos expresiones para $F(y_1)$, debemos tener

$$F(y_1|y_2) f_2(y_2) = \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1$$

o bien

$$F(y_1|y_2) = \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)}\, dt_1.$$

Al integrando de esta expresión lo llamaremos *función de densidad condicional* de Y_1 dado que $Y_2 = y_2$, y lo denotaremos por $f(y_1|y_2)$.

DEFINICIÓN **5.7**

Sean Y_1 y Y_2 variables aleatorias continuas conjuntas con densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente. Para cualquier y_2 tal que $f_2(y_2) > 0$, la densidad condicional de Y_1 dada $Y_2 = y_2$ está dada por

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

y, para cualquier y_1 tal que $f_1(y_1) > 0$, la densidad condicional de Y_2 dada $Y_1 = y_1$ está dada por

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}.$$

Observe que la densidad condicional $f(y_1|y_2)$ es indefinida para toda y_2 tal que $f_2(y_2) = 0$. Del mismo modo, $f(y_2|y_1)$ es indefinida si y_1 es tal que $f_1(y_1) = 0$.

EJEMPLO **5.8** Una máquina automática expendedora de bebidas tiene una cantidad aleatoria Y_2 de bebida en existencia al principio de un día determinado y dosifica una cantidad aleatoria Y_1 durante el día (con cantidades expresadas en galones). La máquina no se reabastece durante el día y, en consecuencia, $Y_1 \le Y_2$. Se ha observado que Y_1 y Y_2 tienen una densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 1/2, & 0 \le y_1 \le y_2 \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Esto es, los puntos (y_1, y_2) están uniformemente distribuidos en el triángulo con las fronteras dadas. Encuentre la densidad condicional de Y_1 dada $Y_2 = y_2$. Evalúe la probabilidad de que se venda menos de $1/2$ galón, dado que la máquina contiene 1.5 galones al empezar el día.

Solución La densidad marginal de Y_2 está dada por

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1.$$

Entonces,

$$f_2(y_2) = \begin{cases} \displaystyle\int_0^{y_2} (1/2)\, dy_1 = (1/2)y_2, & 0 \le y_2 \le 2, \\ \displaystyle\int_{-\infty}^{\infty} 0\, dy_1 = 0, & \text{en cualquier otro punto.} \end{cases}$$

Observe que $f_2(y_2) > 0$ si y sólo si $0 < y_2 \le 2$. Entonces, para cualquier $0 < y_2 \le 2$, usando la Definición 5.7,

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)} = \frac{1/2}{(1/2)(y_2)} = \frac{1}{y_2}, \qquad 0 \le y_1 \le y_2.$$

También, $f(y_1|y_2)$ es indefinida si $y_2 \le 0$ o $y_2 > 2$. La probabilidad de interés es

$$P(Y_1 \le 1/2|Y_2 = 1.5) = \int_{-\infty}^{1/2} f(y_1|y_2 = 1.5)\, dy_1 = \int_0^{1/2} \frac{1}{1.5}\, dy_1 = \frac{1/2}{1.5} = \frac{1}{3}.$$

Si la máquina contiene 2 galones al empezar el día, entonces

$$P(Y_1 \leq 1/2|Y_2 = 2) = \int_0^{1/2} \frac{1}{2}\,dy_1 = \frac{1}{4}.$$

Por tanto, la probabilidad condicional de que $Y_1 \leq 1/2$ dado que $Y_2 = y_2$ cambia de manera apreciable dependiendo de la selección particular de y_2. ■

Ejercicios

5.19 En el Ejercicio 5.1 determinamos que la distribución conjunta de Y_1, el número de contratos otorgados a la empresa A, y Y_2, el número de contratos otorgados a la empresa B, está dada por las entradas en la siguiente tabla.

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

a Encuentre la distribución de probabilidad marginal de Y_1.

b De acuerdo con los resultados vistos en el Capítulo 4, Y_1 tiene una distribución binomial con $n = 2$ y $p = 1/3$. ¿Hay algún conflicto entre este resultado y la respuesta dada en el inciso a?

5.20 Consulte el Ejercicio 5.2.

a Obtenga la distribución de probabilidad marginal para sus ganancias en la apuesta colateral.

b ¿Cuál es la probabilidad de que obtenga tres caras, dado que ganó $1 en la apuesta colateral?

5.21 En el Ejercicio 5.3 determinamos que la distribución de probabilidad conjunta de Y_1, el número de ejecutivos casados, y Y_2, el número de ejecutivos que nunca se han casado, está dada por

$$p(y_1, y_2) = \frac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3 - y_1 - y_2}}{\binom{9}{3}}$$

donde y_1 y y_2 son enteros, $0 \leq y_1 \leq 3$, $0 \leq y_2 \leq 3$ y $1 \leq y_1 + y_2 \leq 3$.

a Encuentre la distribución de probabilidad marginal de Y_1, el número de ejecutivos casados de entre los tres seleccionados para promoción.

b Encuentre $P(Y_1 = 1|Y_2 = 2)$.

c Si con Y_3 denotamos el número de ejecutivos divorciados de entre los tres seleccionados para el cargo, entonces $Y_3 = 3 - Y_1 - Y_2$. Encuentre $P(Y_3 = 1|Y_2 = 1)$.

d Compare la distribución marginal que obtuvo en a con las distribuciones hipergeométricas con $N = 9$, $n = 3$ y $r = 4$ encontradas en la Sección 3.7.

5.22 En el Ejercicio 5.4 nos dieron la siguiente función de probabilidad conjunta para

$$Y_1 = \begin{cases} 0, & \text{si el niño sobrevive,} \\ 1, & \text{si el niño no sobrevive,} \end{cases} \qquad \text{y} \qquad Y_2 = \begin{cases} 0, & \text{si no usaba cinturón,} \\ 1, & \text{si usaba cinturón para adulto,} \\ 2, & \text{si usaba el cinturón del asiento del auto.} \end{cases}$$

	y_1		
y_2	0	1	Total
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
Total	.76	.24	1.00

a Proporcione las funciones de probabilidad marginal para Y_1 y Y_2.

b Proporcione la función de probabilidad condicional para Y_2 dado que $Y_1 = 0$.

c ¿Cuál es la probabilidad de que un niño sobreviva dado que llevaba puesto el cinturón del asiento del auto?

5.23 En el Ejemplo 5.4 y el Ejercicio 5.5 consideramos la densidad conjunta de Y_1, la proporción de la capacidad del tanque que se abastece al principio de la semana, y Y_2, la proporción de la capacidad vendida durante la semana, dada por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la función de densidad marginal para Y_2.

b ¿Para qué valores de y_2 está definida la densidad condicional $f(y_1|y_2)$?

c ¿Cuál es la probabilidad de que más de la mitad del tanque se venda dado que se abastecen tres cuartas partes del tanque?

5.24 En el Ejercicio 5.6 supusimos que si una partícula radiactiva se coloca aleatoriamente en un cuadrado con lados de longitud unitaria, un modelo razonable para la función de densidad conjunta para Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre las funciones de densidad marginal para Y_1 y Y_2.

b ¿Cuál es $P(.3 < Y_1 < .5)$? ¿$P(.3 < Y_2 < .5)$?

c ¿Para qué valores de y_2 está definida la densidad condicional $f(y_1|y_2)$?

d Para cualquier y_2, $0 \le y_2 \le 1$, ¿cuál es la función de densidad condicional de Y_1 dado que $Y_2 = y_2$?

e Encuentre $P(.3 < Y_1 < .5|Y_2 = .3)$.

f Encuentre $P(.3 < Y_1 < .5|Y_2 = .5)$.

g Compare las respuestas que obtuvo en los incisos a, d y e. Para cualquier y_2, $0 \le y_2 \le 1$ ¿cómo se compara $P(.3 < Y_1 < .5)$ con $P(.3 < Y_1 < .5|Y_2 = y_2)$?

5.25 Considere que Y_1 y Y_2 tienen la función de densidad conjunta encontrada primero en el Ejercicio 5.7:

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre las funciones de densidad marginal para Y_1 y Y_2. Identifique estas densidades como una de las estudiadas en el Capítulo 4.

b ¿Cuál es $P(1 < Y_1 < 2.5)$? ¿$P(1 < Y_2 < 2.5)$?

c ¿Para qué valores de y_2 está definida la densidad condicional $f(y_1|y_2)$?

d Para cualquier $y_2 > 0$, ¿cuál es la función de densidad condicional de Y_1 dado que $Y_2 = y_2$?

e Para cualquier $y_1 > 0$, ¿cuál es la función de densidad condicional de Y_1 dado que $Y_1 = y_1$?

f Para cualquier $y_2 > 0$, ¿cómo se compara la función de densidad condicional $f(y_1|y_2)$ que obtuvo en el inciso d con la función de densidad marginal $f_1(y_1)$ hallada en el inciso a?

g ¿Qué implica su respuesta al inciso f acerca de las probabilidades marginales y condicionales de que Y_1 caiga en cualquier intervalo?

5.26 En el Ejercicio 5.8 dedujimos el hecho de que

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta. Encuentre

a las funciones de densidad marginal para Y_1 y Y_2.

b $P(Y_1 \le 1/2|Y_2 \ge 3/4)$.

c la función de densidad condicional de Y_1 dado que $Y_2 = y_2$.

d la función de densidad condicional de Y_2 dado que $Y_1 = y_1$.

e $P(Y_1 \le 3/4|Y_2 = 1/2)$.

5.27 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. Encuentre

a las funciones de densidad marginal para Y_1 y Y_2.

b $P(Y_2 \le 1/2|Y_1 \le 3/4)$.

c la función de densidad condicional de Y_1 dado que $Y_2 = y_2$.

d la función de densidad condicional de Y_2 dado que $Y_1 = y_1$.

e $P(Y_2 \ge 3/4|Y_1 = 1/2)$.

5.28 En el Ejercicio 5.10 demostramos que

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 2, 0 \le y_2 \le 1, 2y_2 \le y_1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida para Y_1, la cantidad de contaminante por muestra recolectada arriba de la chimenea que no tenía el aparato limpiador, y para Y_2, la cantidad recolectada arriba de la chimenea con el aparato limpiador.

a Si consideramos la chimenea con el limpiador instalado, encuentre la probabilidad de que la cantidad de contaminante en una muestra determinada sea mayor que .5.

b Dado que se observa que la cantidad de contaminante en una muestra tomada arriba de la chimenea con el limpiador es 0.5, encuentre la probabilidad de que la cantidad de contaminante exceda de 1.5 arriba de la otra chimenea (la que no tiene limpiador).

5.29 Consulte el Ejercicio 5.11. Encuentre

a las funciones de densidad marginal para Y_1 y Y_2.

b $P(Y_2 > 1/2|Y_1 = 1/4)$.

5.30 En el Ejercicio 5.12 nos dieron la siguiente función de densidad de probabilidad conjunta para las variables aleatorias Y_1 y Y_2, que fueron las proporciones de dos componentes en una muestra tomada de una mezcla de insecticida:

$$f(y_1, y_2) = \begin{cases} 2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, 0 \le y_1 + y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $P(Y_1 \ge 1/2 | Y_2 \le 1/4)$.
b Encuentre $P(Y_1 \ge 1/2 | Y_2 = 1/4)$.

5.31 En el Ejercicio 5.13 la función de densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} 30 y_1 y_2^2, & y_1 - 1 \le y_2 \le 1 - y_1, 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que la densidad marginal de Y_1 es una densidad beta con $\alpha = 2$ y $\beta = 4$.
b Obtenga la densidad marginal de Y_2.
c Obtenga la densidad condicional de Y_2 dado que $Y_1 = y_1$.
d Encuentre $P(Y_2 > 0 | Y_1 = .75)$.

5.32 Suponga que las variables aleatorias Y_1 y Y_2 tienen función de densidad de probabilidad conjunta, $f(y_1, y_2)$, dada por (vea el Ejercicio 5.14)

$$f(y_1, y_2) = \begin{cases} 6 y_1^2 y_2, & 0 \le y_1 \le y_2, y_1 + y_2 \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que la densidad marginal de Y_1 es una densidad beta con $\alpha = 3$ y $\beta = 2$.
b Deduzca la densidad marginal de Y_2.
c Deduzca la densidad condicional de Y_2 dado que $Y_1 = y_1$.
d Encuentre $P(Y_2 < 1.1 | Y_1 = .60)$.

5.33 Suponga que Y_1 es el tiempo total entre la llegada de un cliente a la tienda y su salida desde la ventanilla de servicio, Y_2 es el tiempo empleado en la fila de espera antes de llegar a la ventanilla y la densidad conjunta de estas variables (como se da en el Ejercicio 5.15) es

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \le y_2 \le y_1 \le \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre las funciones de densidad marginal para Y_1 y Y_2.
b ¿Cuál es la función de densidad condicional de Y_1 dado que $Y_2 = y_2$? Asegúrese de especificar los valores de y_2 para los cuales está definida esta densidad condicional.
c ¿Cuál es la función de densidad condicional de Y_2 dado que $Y_1 = y_1$? Asegúrese de especificar los valores de y_1 para los cuales está definida esta densidad condicional.
d ¿La función de densidad condicional $f(y_1|y_2)$ que obtuvo en el inciso b es la misma que la función de densidad marginal $f_1(y_1)$ hallada en el inciso a?
e ¿Qué implica su respuesta al inciso d acerca de las probabilidades marginal y condicional de que Y_1 caiga en cualquier intervalo?

5.34 Si Y_1 está uniformemente distribuida en el intervalo (0, 1) y, para $0 < y_1 < 1$,

$$f(y_2 | y_1) = \begin{cases} 1/y_1, & 0 \le y_2 \le y_1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Cuál es el "nombre" de la distribución condicional de Y_2 dado que $Y_1 = y_1$?
b Encuentre la función de densidad conjunta de Y_1 y Y_2.
c Encuentre la función de densidad marginal para Y_2.

5.35 Consulte el Ejercicio 5.33. Si transcurren dos minutos entre la llegada de un cliente a una tienda y su salida de la ventanilla de servicio, encuentre la probabilidad de que espere en la fila menos de un minuto para llegar a la ventanilla.

5.36 En el Ejercicio 5.16, Y_1 y Y_2 denotaron las proporciones de tiempo durante las cuales los empleados I y II en realidad ejecutaron sus tareas asignadas durante un día de trabajo. La densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre las funciones de densidad marginales para Y_1 y Y_2.

b Encuentre $P(Y_1 \ge 1/2|Y_2 \ge 1/2)$.

c Si el empleado II pasa exactamente 50% del día trabajando en sus tareas asignadas, encuentre la probabilidad de que el empleado I pase más de 75% del día trabajando en tareas similares.

5.37 En el Ejercicio 5.18, Y_1 y Y_2 denotaron el tiempo de vida útil, en cientos de horas, para componentes de los tipos I y II, respectivamente, en un sistema electrónico. La densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la probabilidad de que el componente tipo II tenga una vida útil de más de 200 horas.

5.38 Denote con Y_1 el peso (en toneladas) de un artículo a granel que un proveedor tiene en existencia al principio de una semana y suponga que Y_1 tiene una distribución uniforme en el intervalo $0 \le y_1 \le 1$. Denote con Y_2 la cantidad (en peso) de este artículo vendido por el proveedor durante la semana y suponga que Y_2 tiene una distribución uniforme en el intervalo $0 \le y_2 \le y_1$, donde y_1 es un valor específico de Y_1.

a Encuentre la función de densidad conjunta para Y_1 y Y_2.

b Si el proveedor tiene en existencia media tonelada del artículo, ¿cuál es la probabilidad de que venda más de un cuarto de tonelada?

c Si se sabe que el proveedor vendió un cuarto de tonelada del artículo, ¿cuál es la probabilidad de que hubiera tenido en existencia más de media tonelada?

***5.39** Suponga que Y_1 y Y_2 son variables aleatorias independientes con distribución de Poisson, con medias λ_1 y λ_2, respectivamente. Sea $W = Y_1 + Y_2$. En el Capítulo 6 demostraremos que W tiene una distribución de Poisson con media $\lambda_1 + \lambda_2$. Use este resultado para demostrar que la distribución condicional de Y_1, dado que $W = w$, es una distribución binomial con $n = w$ y $p = \lambda_1/(\lambda_1 + \lambda_2)$.[1]

***5.40** Suponga que Y_1 y Y_2 son variables aleatorias independientes con distribución binomial basadas en muestras de tamaños n_1 y n_2, respectivamente. Suponga que $p_1 = p_2 = p$. Esto es, la probabilidad de "éxito" es la misma para las dos variables aleatorias. Sea $W = Y_1 + Y_2$. En el Capítulo 6 demostraremos que W tiene una distribución binomial con probabilidad de éxito p y tamaño muestral $n_1 + n_2$. Use este resultado para demostrar que la distribución condicional de Y_1, dado que $W = w$, es una distribución hipergeométrica con $N = n_1 + n_2$, $n = w$ y $r = n_1$.

***5.41** Un plan de control de calidad exige seleccionar aleatoriamente tres artículos provenientes de la producción diaria (supuestamente grande) de cierta máquina y observar el número de artículos defectuosos. No obstante, la proporción p de artículos defectuosos producidos por la máquina varía de un día a otro y se supone que tiene una distribución uniforme en el intervalo (0, 1). Para un día escogido aleatoriamente, encuentre la probabilidad incondicional de que se observen exactamente dos artículos defectuosos en la muestra.

1. Los ejercicios precedidos por un asterisco son opcionales.

***5.42** Se sabe que el número de defectos por yarda Y para cierta tela tiene una distribución de Poisson, con parámetro λ. No obstante, λ es una variable aleatoria con función de densidad de probabilidad dada por

$$f(\lambda) = \begin{cases} e^{-\lambda}, & \lambda \geq 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de probabilidad incondicional para Y.

5.4 Variables aleatorias independientes

En el Ejemplo 5.8 vimos dos variables aleatorias dependientes, para las cuales las probabilidades asociadas con Y_1 dependían del valor observado de Y_2. En el Ejercicio 5.24 (y algunos otros) éste no fue el caso: las probabilidades asociadas con Y_1 eran iguales, cualquiera que fuera el valor observado de Y_2. Ahora presentamos una definición formal de *independencia* de variables aleatorias.

Dos eventos A y B son independientes si $P(A \cap B) = P(A) \times P(B)$. Cuando estudiemos variables aleatorias, si $a < b$ y $c < d$ es frecuente que nos interesemos en eventos del tipo $(a < Y_1 \leq b) \cap (c < Y_2 \leq d)$. Por consistencia con la definición anterior de eventos independientes, si Y_1 y Y_2 son independientes, nos gustaría tener

$$P(a < Y_1 \leq b,\ c < Y_2 \leq d) = P(a < Y_1 \leq b) \times P(c < Y_2 \leq d)$$

para cualquier elección de números reales $a < b$ y $c < d$. Esto es, si Y_1 y Y_2 son independientes, la probabilidad conjunta se puede escribir como el producto de las probabilidades marginales. Esta propiedad se satisface si Y_1 y Y_2 son independientes en el sentido detallado en la siguiente definición.

DEFINICIÓN **5.8**

Sea Y_1 que tiene una función de distribución $F_1(y_1)$ y sea Y_2 que tiene una función de distribución $F_2(y_2)$, y $F(y_1, y_2)$ es la función de distribución conjunta de Y_1 y Y_2. Entonces se dice que Y_1 y Y_2 son *independientes* si y sólo si

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

para todo par de números reales $(y_1,\ y_2)$.

Si Y_1 y Y_2 no son independientes, se dice que son *dependientes.*

Por lo general es cómodo establecer la presencia o ausencia de independencia, por medio del resultado del siguiente teorema. Se omite la demostración; vea "Bibliografía y lecturas adicionales" al final del capítulo.

TEOREMA **5.4**

Si Y_1 y Y_2 son variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$ y funciones de probabilidad marginal $p_1(y_1)$ y $p_2(y_2)$, respectivamente, entonces Y_1 y Y_2 son independientes si y sólo si

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

para todos los pares de números reales $(y_1,\ y_2)$.

Si Y_1 y Y_2 son variables aleatorias continuas con función de densidad conjunta $f_1(y_1, y_2)$ y funciones de densidad marginal $f_1(y_1)$ y $f_2(y_2)$, respectivamente, entonces Y_1 y Y_2 son independientes si y sólo si

$$f(y_1, y_2) = f_1(y_1) f_2(y_2)$$

para todos los pares de números reales (y_1, y_2).

A continuación ilustramos el concepto de independencia con algunos ejemplos.

EJEMPLO **5.9** Para el problema de tirar un dado de la Sección 5.2, demuestre que Y_1 y Y_2 son independientes.

Solución En este problema a cada uno de los 36 puntos muestrales se le dio probabilidad $1/36$. Considere, por ejemplo, el punto $(1, 2)$. Sabemos que $p(1, 2) = 1/36$. También, $p_1(1) = P(Y_1 = 1) = 1/6$ y $p_2(2) = P(Y_2 = 2) = 1/6$. Por tanto,

$$p(1, 2) = p_1(1)\, p_2(2).$$

Lo mismo es cierto para *todos los demás valores* de y_1 y y_2, de lo cual se deduce que Y_1 y Y_2 son independientes. ■

EJEMPLO **5.10** Consulte el Ejemplo 5.5. ¿El número de republicanos en la muestra es independiente del número de demócratas? (¿Es Y_1 independiente de Y_2?)

Solución La independencia de variables aleatorias discretas requiere que $p(y_1, y_2) = p_1(y_1)\, p_2(y_2)$ para toda selección (y_1, y_2). Entonces, si esta igualdad es violada para cualquier par de valores (y_1, y_2), las variables aleatorias son dependientes. Al observar la esquina superior izquierda de la Tabla 5.2, veremos que

$$P(0, 0) = 0.$$

Pero $p_1(0) = 3/15$ y $p_2(0) = 6/15$. En consecuencia,

$$p(0, 0) \neq p_1(0)\, p_2(0),$$

de modo que Y_1 y Y_2 son dependientes. ■

EJEMPLO **5.11** Sea

$$f(y_1, y_2) = \begin{cases} 6y_1 y_2^2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que Y_1 y Y_2 son independientes.

Solución Tenemos

$$f_1(y_1) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} f(y_1, y_2)\,dy_2 = \int_0^1 6y_1y_2^2\,dy_2 = 6y_1\left(\frac{y_2^3}{3}\bigg]_0^1\right) & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad = 2y_1, & 0 \le y_1 \le 1, \\ \displaystyle\int_{-\infty}^{\infty} f(y_1, y_2)\,dy_2 = \int_{-\infty}^{\infty} 0\,dy_1 = 0, & \text{en cualquier otro punto.} \end{cases}$$

Del mismo modo,

$$f_2(y_2) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} f(y_1, y_2)\,dy_1 = \int_0^1 6y_1y_2^2\,dy_1 = 3y_2^2, & 0 \le y_2 \le 1, \\ \displaystyle\int_{-\infty}^{\infty} f(y_1, y_2)\,dy_1 = \int_{-\infty}^{\infty} 0\,dy_1 = 0, & \text{en cualquier otro punto.} \end{cases}$$

En consecuencia,

$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

para todos los números reales (y_1, y_2) y, por tanto, Y_1 y Y_2 son independientes. ■

EJEMPLO **5.12** Sea

$$f(y_1, y_2) = \begin{cases} 2, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que Y_1 y Y_2 son dependientes.

Solución Vemos que $f(y_1, y_2) = 2$ sobre la región sombreada que se ve en la Figura 5.7. Por tanto,

$$f_1(y_1) = \begin{cases} \displaystyle\int_0^{y_1} 2\,dy_2 = 2y_2\bigg]_0^{y_1} = 2y_1, & 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

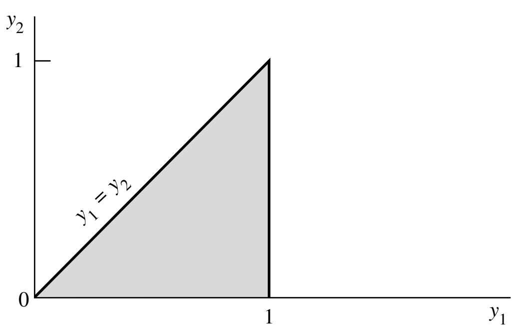


FIGURA **5.7**
Región sobre la cual $f(y_1, y_2)$ es positiva, Ejemplo 5.12

Del mismo modo,

$$f_2(y_2) = \begin{cases} \displaystyle\int_{y_2}^{1} 2\,dy_1 = 2y_1\Big]_{y_2}^{1} = 2(1 - y_2), & 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por tanto,

$$f(y_1, y_2) \neq f_1(y_1) f_2(y_2)$$

para algún par de números reales $(y_1,\ y_2)$ y, por tanto, Y_1 y Y_2 son dependientes. ■

Observará una diferencia distinta en los límites de integración empleados para hallar las funciones de densidad marginal obtenidas en los Ejemplos 5.11 y 5.12. Los límites de integración para y_2, comprendidos en hallar la densidad marginal de Y_1 en el Ejemplo 5.12, dependían de y_1. En contraste, los límites de integración fueron constantes cuando determinamos las funciones de densidad marginal del Ejemplo 5.11. Si los límites de integración son constantes, el siguiente teorema proporciona una forma fácil de demostrar la independencia de dos variables aleatorias.

TEOREMA 5.5 Sean Y_1 y Y_2 que tienen una densidad conjunta $f(y_1, y_2)$ que es positiva si y sólo si $a \le y_1 \le b$ y $c \le y_2 \le d$, para constantes a, b, c y d; y $f(y_1, y_2) = 0$ en otro caso. Entonces Y_1 y Y_2 son variables aleatorias independientes si y sólo si

$$f(y_1, y_2) = g(y_1)\,h(y_2)$$

donde $g(y_1)$ es una función no negativa de y_1 solamente y $h(y_2)$ es una función no negativa de y_2 solamente.

La demostración de este teorema se omite. (Vea "Bibliografía y lecturas adicionales" al final del capítulo.) El beneficio clave del resultado dado en el Teorema 5.5 es que en realidad no necesitamos obtener las densidades marginales. De hecho, las funciones $g(y_1)$ y $h(y_2)$ no necesitan ser funciones de densidad (aun cuando sean múltiplos constantes de las densidades marginales, deberíamos tomarnos la molestia de determinar éstas).

EJEMPLO 5.13 Sean Y_1 y Y_2 que tienen una densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Y_1 y Y_2 son variables independientes?

Solución Observe que $f(y_1, y_2)$ es positiva si y sólo si $0 \le y_1 \le 1$ y $0 \le y_2 \le 1$. Además, $f(y_1, y_2) = g(y_1)\,h(y_2)$, donde

$$g(y_1) = \begin{cases} y_1, & 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases} \quad \text{y} \quad h(y_2) = \begin{cases} 2, & 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por tanto, Y_1 y Y_2 son variables aleatorias independientes. Observe que $g(y_1)$ y $h(y_2)$, como aquí se definen, *no son* funciones de densidad, aun cuando $2g(y_1)$ y $h(y_2)/2$ sean densidades. ■

EJEMPLO **5.14** Consulte el Ejemplo 5.4. ¿Y_1, la cantidad en existencia, es independiente de Y_2, la cantidad vendida?

Solución Como la función de densidad es positiva si y sólo si $0 \le y_2 \le y_1 \le 1$, no existen *constantes a, b, c* y d tales que la densidad sea positiva en la región $a \le y_1 \le b$, $c \le y_2 \le d$. Entonces, el Teorema 5.5 no se puede aplicar. No obstante, se puede demostrar que Y_1 y Y_2 son variables aleatorias dependientes porque la densidad conjunta no es el producto de las densidades marginales. ■

Las definiciones 5.8 fácilmente se pueden generalizar a n dimensiones. Suponga que tenemos n variables aleatorias, $Y_1, \ldots, Y_n$, donde Y_i tiene función de distribución $F_i(y_i)$, para $i = 1, 2, \ldots, n$; y donde $Y_1, Y_2, \ldots, Y_n$ tienen función de distribución conjunta $F(y_1, y_2, \ldots, y_n)$. Entonces $Y_1, Y_2, \ldots, Y_n$ son independientes si y sólo si

$$F(y_1, y_2, \ldots, y_n) = F_1(y_1) \cdots F_n(y_n)$$

para todos los números reales $y_1, y_2, \ldots, y_n$, con las formas equivalentes obvias para los casos discretos y continuos.

Ejercicios

5.43 Sean Y_1 y Y_2 que tienen una función de densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente. Demuestre que Y_1 y Y_2 son independientes si y sólo si $f(y_1|y_2) = f_1(y_1)$ para todos los valores de y_1 y para toda y_2 tal que $f_2(y_2) > 0$. Un argumento completamente análogo establece que Y_1 y Y_2 son independientes si y sólo si $f(y_2|y_1) = f_2(y_2)$ para todos los valores de y_2 y para toda y_1 tal que $f_1(y_1) > 0$.

5.44 Demuestre que los resultados del Ejercicio 5.43 también se cumplen para variables aleatorias discretas.

5.45 En el Ejercicio 5.1 determinamos que la distribución conjunta de Y_1, el número de contratos concedidos a la empresa A, y Y_2, el número de contratos concedidos a la empresa B, está dada por las entradas de la tabla siguiente.

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

La función de probabilidad marginal de Y_1 se determinó en el Ejercicio 5.19 como binomial con $n = 2$ y $p = 1/3$. ¿Y_1 y Y_2 son independientes? ¿Por qué?

5.46 Consulte el Ejercicio 5.2. El número de caras en tres tiros de moneda está binomialmente distribuido con $n = 3$, $p = 1/2$. ¿Son independientes el número total de caras y sus ganancias en la apuesta colateral? [Examine su respuesta al Ejercicio 5.20(b).]

5.47 En el Ejercicio 5.3 determinamos que la distribución de probabilidad conjunta de Y_1, el número de ejecutivos casados y Y_2, el número de ejecutivos que no se han casado, está dada por

$$p(y_1, y_2) = \frac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3 - y_1 - y_2}}{\binom{9}{3}},$$

donde y_1 y y_2 son enteros, $0 \leq y_1 \leq 3$, $0 \leq y_2 \leq 3$ y $1 \leq y_1 + y_2 \leq 3$. ¿Y_1 y Y_2 son independientes? (Recuerde su respuesta al Ejercicio 5.21.)

5.48 En el Ejercicio 5.4 se determinó la siguiente función de probabilidad conjunta para

$$Y_1 = \begin{cases} 0, & \text{si el niño sobrevive,} \\ 1, & \text{si no sobrevive,} \end{cases} \quad \text{y} \quad Y_2 = \begin{cases} 0, & \text{si no usaba cinturón,} \\ 1, & \text{si usaba cinturón para adulto,} \\ 2, & \text{si usaba el cinturón del asiento del auto.} \end{cases}$$

	y_1		
y_2	0	1	Total
0	.38	.17	.55
1	.14	.02	.16
2	.24	.05	.29
Total	.76	.24	1.00

¿Y_1 y Y_2 son independientes? ¿Por qué sí o por qué no?

5.49 En el Ejemplo 5.4 y en el Ejercicio 5.5 consideramos la densidad conjunta de Y_1, la proporción de la capacidad del tanque que ha sido abastecido al principio de la semana y Y_2, la proporción de la capacidad vendida durante la semana, dadas por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \leq y_2 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que Y_1 y Y_2 son dependientes.

5.50 En el Ejercicio 5.6 supusimos que si una partícula radiactiva se coloca en un cuadrado con lados de longitud unitaria, un modelo razonable para la función de densidad conjunta para Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} 1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Y_1 y Y_2 son independientes?

b ¿El resultado del inciso a explica los resultados obtenidos en el Ejercicio 5.24 d - f? ¿Por qué?

5.51 En el Ejercicio 5.7 consideramos Y_1 y Y_2 con función de densidad conjunta

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Y_1 y Y_2 son independientes?

b ¿El resultado del inciso a explica los resultados obtenidos en el Ejercicio 5.25 d - f? ¿Por qué?

5.52 En el Ejercicio 5.8 dedujimos que

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. ¿Y_1 y Y_2 son independientes?

5.53 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. ¿Y_1 y Y_2 son independientes?

5.54 En el Ejercicio 5.10 demostramos que

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 2, 0 \le y_2 \le 1, 2y_2 \le y_1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida para Y_1, la cantidad de contaminante por muestra recolectada arriba de la chimenea que no está equipada con aparato limpiador, y Y_2, la cantidad recolectada arriba de la chimenea con limpiador. ¿Son independientes las cantidades de contaminantes por muestra recolectadas con y sin el aparato limpiador?

5.55 Suponga que, como en el Ejercicio 5.11, Y_1 y Y_2 están uniformemente distribuidas sobre el triángulo sombreado en el diagrama adjunto. ¿Y_1 y Y_2 son independientes?

5.56 En el Ejercicio 5.12 se determinó la siguiente función de densidad de probabilidad conjunta para las variables aleatorias Y_1 y Y_2, que eran las proporciones de dos componentes en una muestra de una mezcla de insecticida:

$$f(y_1, y_2) = \begin{cases} 2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, 0 \le y_1 + y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Y_1 y Y_2 son independientes?

5.57 En los Ejercicios 5.13 y 5.31, la función de densidad conjunta de Y_1 y Y_2 fue dada por

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \le y_2 \le 1 - y_1, 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Las variables aleatorias Y_1 y Y_2 son independientes?

5.58 Suponga que las variables aleatorias Y_1 y Y_2 tienen función de densidad de probabilidad conjunta, $f(y_1, y_2)$, dada por (vea Ejercicios 5.14 y 5.32)

$$f(y_1, y_2) = \begin{cases} 6y_1^2\, y_2, & 0 \le y_1 \le y_2, y_1 + y_2 \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que Y_1 y Y_2 son variables aleatorias dependientes.

5.59 Si Y_1 es el tiempo total entre la llegada de un cliente a una tienda y su salida de la ventanilla de servicio, y si Y_2 es el tiempo empleado en la fila de espera antes de llegar a la ventanilla, la densidad conjunta de estas variables, de acuerdo con el Ejercicio 5.15, es

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \le y_2 \le y_1 \le \infty \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Y_1 y Y_2 son independientes?

5.60 En el Ejercicio 5.16, Y_1 y Y_2 denotaron las proporciones de tiempo que los empleados I y II pasaron realmente trabajando en sus tareas asignadas durante un día hábil. La densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Y_1 y Y_2 son independientes?

5.61 En el Ejercicio 5.18, Y_1 y Y_2 denotaron las duraciones de vida útil, en cientos de horas, para componentes de los tipos I y II, respectivamente, en un sistema electrónico. La densidad conjunta de Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} (1/8) y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿Y_1 y Y_2 son independientes?

5.62 Suponga que la probabilidad de que aparezca una cara cuando una moneda se lanza al aire es p y que la probabilidad de que aparezca una cruz es $q = 1 - p$. La persona A lanza la moneda hasta que aparece la primera cara y se detiene. La persona B hace lo mismo. Los resultados obtenidos por las personas A y B se supone que son independientes. ¿Cuál es la probabilidad de que A y B se detengan en exactamente el mismo número de tiros?

5.63 Sean Y_1 y Y_2 variables aleatorias independientes distribuidas exponencialmente, cada una con media 1. Encuentre $P(Y_1 > Y_2 | Y_1 < 2Y_2)$.

5.64 Sean Y_1 y Y_2 variables aleatorias independientes que están distribuidas uniformemente en el intervalo (0, 1). Encuentre $P(Y_1 < 2Y_2 | Y_1 < 3Y_2)$.

***5.65** Suponga que, para $-1 \le \alpha \le 1$, la función de densidad de probabilidad de (Y_1, Y_2) está dada por

$$f(y_1, y_2) = \begin{cases} [1 - \alpha\{(1 - 2e^{-y_1})(1 - 2e^{-y_2})\}]e^{-y_1-y_2}, & 0 \le y_1, 0 \le y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que la distribución marginal de Y_1 es exponencial con media 1.

b ¿Cuál es la distribución marginal de Y_2?

c Demuestre que Y_1 y Y_2 son independientes si y sólo si $\alpha = 0$.

Observe que estos resultados implican que hay un número infinito de densidades conjuntas tales que ambas densidades marginales son exponenciales con media 1.

***5.66** Sean $F_1(y_1)$ y $F_2(y_2)$ dos funciones de distribución. Para cualquier α, $-1 \le \alpha \le 1$, considere Y_1 y Y_2 con función de distribución conjunta

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)[1 - \alpha\{1 - F_1(y_1)\}\{1 - F_2(y_2)\}].$$

a ¿Cuál es $F(y_1, \infty)$, la función de distribución marginal de Y_1? [*Sugerencia:* ¿cuál es $F_2(\infty)$?]

b ¿Cuál es la función de distribución marginal de Y_2?

c Si $\alpha = 0$ ¿por qué son independientes Y_1 y Y_2?

d ¿Y_1 y Y_2 son independientes si $\alpha \ne 0$? ¿Por qué?

Observe que esta construcción se puede usar para producir un número infinito de funciones de distribución conjunta que tengan las mismas funciones de distribución marginal.

5.67 En la Sección 5.2 dijimos que si Y_1 y Y_2 tienen una función de distribución acumulativa conjunta $F(y_1, y_2)$, entonces para cualquier $a < b$ y $c < d$

$$P(a < Y_1 \le b, c < Y_2 \le d) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Si Y_1 y Y_2 son independientes, demuestre que

$$P(a < Y_1 \leq b, c < Y_2 \leq d) = P(a < Y_1 \leq b) \times P(c < Y_2 \leq d).$$

[*Sugerencia:* exprese $P(a < Y_1 \leq b)$ en términos de $F_1(\cdot)$.]

5.68 Un supermercado tiene dos clientes esperando para pagar sus compras en la caja I y un cliente esperando pagar en la caja II. Denote con Y_1 y Y_2 los números de clientes que gastan más de \$50 en abarrotes en las cajas respectivas. Suponga que Y_1 y Y_2 son variables aleatorias binomiales independientes, la probabilidad de que un cliente en la caja I gaste más de \$50 es .2 y la probabilidad de que un cliente en la caja II gaste más de \$50 es igual a .3. Encuentre

a la distribución de probabilidad conjunta para Y_1 y Y_2.
b la probabilidad de que no más de uno de los tres clientes gaste más de \$50.

5.69 La vida útil Y para cierto tipo de fusibles está modelada por la distribución exponencial, con

$$f(y) = \begin{cases} (1/3)e^{-y/3}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(Las mediciones son en cientos de horas.)

a Si dos de esos fusibles tienen vidas útiles independientes Y_1 y Y_2, encuentre la función de densidad de probabilidad conjunta para Y_1 y Y_2.
b Un fusible en el inciso a está en un sistema primario y el otro está en el sistema de respaldo que entra en uso sólo si falla el sistema primario. La vida útil efectiva total de los dos fusibles es entonces $Y_1 + Y_2$. Encuentre $P(Y_1 + Y_2 \leq 1)$.

5.70 Un autobús llega a una parada en un tiempo uniformemente distribuido en el intervalo 0 a 1 hora. Un pasajero también llega a la parada en un tiempo uniformemente distribuido en el intervalo de 0 a 1 hora. Suponga que los tiempos de llegada del autobús y el pasajero son independientes entre sí y que el pasajero va a esperar hasta $1/4$ de hora a que llegue el autobús. ¿Cuál es la probabilidad de que el pasajero aborde el autobús? [*Sugerencia:* denote con Y_1 la hora de llegada del autobús y con Y_2 la hora de llegada del pasajero; determine la densidad conjunta de Y_1 y Y_2 y encuentre $P(Y_2 \leq Y_1 \leq Y_2 + 1/4)$.]

5.71 Dos llamadas telefónicas entran en un conmutador en tiempos aleatorios en un periodo fijo de una hora. Suponga que las llamadas se hacen independientemente una de la otra. ¿Cuál es la probabilidad de que las llamadas se hagan

a en la primera media hora?,
b a no más de cinco minutos entre sí?

5.5 El valor esperado de una función de variables aleatorias

Para justificar la siguiente definición sólo se necesita construir el equivalente multivariante del caso univariante.

DEFINICIÓN **5.9**

Sea $g(Y_1, Y_2, \ldots, Y_k)$ una función de las variables aleatorias discretas, $Y_1, Y_2, \ldots, Y_k$, que tienen función de probabilidad $p(y_1, y_2, \ldots, y_k)$. Entonces el *valor esperado* de $g(Y_1, Y_2, \ldots, Y_k)$ es

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \sum_{\text{toda } y_k} \cdots \sum_{\text{toda } y_2} \sum_{\text{toda } y_1} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k).$$

Si $Y_1, Y_2, \ldots, Y_k$ son variables aleatorias continuas con función de densidad conjunta $f(y_1, y_2, \ldots, y_k)$, entonces[2]

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) \times f(y_1, y_2, \ldots, y_k)\, dy_1\, dy_2 \ldots dy_k.$$

EJEMPLO **5.15** Considere que Y_1 y Y_2 tienen una densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $E(Y_1 Y_2)$.

Solución De la Definición 5.9 obtenemos

$$\begin{aligned} E(Y_1 Y_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f(y_1, y_2)\, dy_1\, dy_2 = \int_0^1 \int_0^1 y_1 y_2 (2y_1)\, dy_1\, dy_2 \\ &= \int_0^1 y_2 \left(\left. \frac{2y_1^3}{3} \right]_0^1 \right) dy_2 = \int_0^1 \left(\frac{2}{3} \right) y_2\, dy_2 = \left. \frac{2}{3} \frac{y_2^2}{2} \right]_0^1 = \frac{1}{3}. \end{aligned}$$

■

Demostraremos que la Definición 5.9 es consistente con la Definición 4.5, en la que definimos el valor esperado de una variable aleatoria univariante. Considere dos variables aleatorias Y_1 y Y_2 con función de densidad $f(y_1, y_2)$. Deseamos hallar el valor esperado de $g(Y_1, Y_2) = Y_1$.

De la Definición 5.9 tenemos

$$\begin{aligned} E(Y_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_2\, dy_1 \\ &= \int_{-\infty}^{\infty} y_1 \left[\int_{-\infty}^{\infty} f(y_1, y_2)\, dy_2 \right] dy_1. \end{aligned}$$

La cantidad dentro de paréntesis rectangulares, por definición, es la función de densidad marginal para Y_1. Por tanto, obtenemos

$$E(Y_1) = \int_{-\infty}^{\infty} y_1 f_1(y_1)\, dy_1,$$

que está acorde con la Definición 4.5.

2. De nuevo, decimos que existen valores esperados si $\sum \cdots \sum |g(y_1, y_2, \ldots, y_n)| p(y_1, y_2, \ldots, y_k)$ o si $\int \cdots \int |g(y_1, y_2, \ldots, y_n)| f(y_1, y_2, \ldots, y_k)\, dy_1 \ldots dy_k$ es finita.

EJEMPLO **5.16** Considere que Y_1 y Y_2 que tienen una densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre el valor esperado de Y_1.

Solución

$$\begin{aligned} E(Y_1) &= \int_0^1 \int_0^1 y_1(2y_1)\, dy_1\, dy_2 \\ &= \int_0^1 \left(\frac{2y_1^3}{3} \bigg]_0^1 \right) dy_2 = \int_0^1 \frac{2}{3}\, dy_2 = \frac{2}{3} y_2 \bigg]_0^1 = \frac{2}{3}. \end{aligned}$$

Consulte la Figura 5.6 y calcule el valor esperado de Y_1. El valor $E(Y_1) = 2/3$ parece ser bastante razonable. ■

EJEMPLO **5.17** En la Figura 5.6 el valor medio de Y_2 parece ser igual a .5. Confirmemos este cálculo visual. Encuentre $E(Y_2)$.

Solución

$$\begin{aligned} E(Y_2) &= \int_0^1 \int_0^1 y_2(2y_1)\, dy_1\, dy_2 = \int_0^1 y_2 \left(\frac{2y_1^2}{2} \bigg]_0^1 \right) dy_2 \\ &= \int_0^1 y_2\, dy_2 = \frac{y_2^2}{2} \bigg]_0^1 = \frac{1}{2}. \end{aligned}$$

■

EJEMPLO **5.18** Sean Y_1 y Y_2 variables aleatorias con función de densidad

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $V(Y_1)$.

Solución La densidad marginal para Y_1 obtenida en el Ejemplo 5.6 es

$$f_1(y_1) = \begin{cases} 2y_1, & 0 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Entonces $V(Y_1) = E\left(Y_1^2\right) - [E(Y_1)]^2$, y

$$E\left(Y_1^k\right) = \int_{-\infty}^{\infty} y_1^k f_1(y_1)\, dy_1 = \int_0^1 y_1^k (2y_1)\, dy_1 = \frac{2y_1^{k+2}}{k+2} \bigg]_0^1 = \frac{2}{k+2}.$$

Si hacemos $k = 1$ y $k = 2$, se deduce que $E(Y_1)$ y $E\left(Y_1^2\right)$ son 2/3 y 1/2, respectivamente. Entonces $V(Y_1) = E\left(Y_1^2\right) - [E(Y_1)]^2 = 1/2 - (2/3)^2 = 1/18$. ■

EJEMPLO **5.19** Del proceso para producir una sustancia química industrial se obtiene un producto que contiene dos tipos de impurezas. Para una muestra específica proveniente de este proceso, denotemos con Y_1 la proporción de impurezas en la muestra y con Y_2 la proporción de impurezas tipo I entre todas las impurezas halladas. Suponga que la distribución conjunta de Y_1 y Y_2 puede ser modelada con la siguiente función de densidad de probabilidad:

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre el valor esperado de la proporción de impurezas tipo I de la muestra.

Solución Como Y_1 es la proporción de impurezas en la muestra y Y_2 es la proporción de impurezas tipo I entre las impurezas muestrales, se deduce que Y_1Y_2 es la proporción de impurezas tipo I en toda la muestra. Entonces, buscamos hallar $E(Y_1Y_2)$:

$$\begin{aligned} E(Y_1Y_2) &= \int_0^1 \int_0^1 2y_1y_2(1 - y_1)\,dy_2\,dy_1 = 2\int_0^1 y_1(1 - y_1)\left(\frac{1}{2}\right) dy_1 \\ &= \int_0^1 \left(y_1 - y_1^2\right) dy_1 = \left(\frac{y_1^2}{2} - \frac{y_1^3}{3}\right)\Bigg]_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}. \end{aligned}$$

Por tanto, esperaríamos que 1/6 de la muestra estuviera formado por impurezas tipo I. ■

5.6 Teoremas especiales

Los teoremas que facilitan el cálculo del valor esperado de una constante, el valor esperado de una constante por una función de variables aleatorias y el valor esperado de la suma de funciones de variables aleatorias son semejantes a los del caso univariante.

TEOREMA **5.6** Sea c una constante. Entonces

$$E(c) = c.$$

TEOREMA **5.7** Sea $g(Y_1, Y_2)$ una función de las variables aleatorias Y_1 y Y_2 y sea c una constante. Entonces

$$E[cg(Y_1, Y_2)] = cE[g(Y_1, Y_2)].$$

TEOREMA 5.8 Sean Y_1 y Y_2 variables aleatorias y $g_1(Y_1, Y_2)$, $g_2(Y_1, Y_2), \ldots, g_k(Y_1, Y_2)$ funciones de Y_1 y Y_2. Entonces

$$\begin{aligned} &E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) + \cdots + g_k(Y_1, Y_2)] \\ &\quad = E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + \cdots + E[g_k(Y_1, Y_2)]. \end{aligned}$$

Las demostraciones de estos tres teoremas son análogas a los casos univariantes estudiados en los Capítulos 3 y 4.

EJEMPLO 5.20 Consulte el Ejemplo 5.4. La variable aleatoria $Y_1 - Y_2$ denota la cantidad proporcional de gasolina remanente al final de la semana. Encuentre $E(Y_1 - Y_2)$.

Solución Empleando el Teorema 5.8 con $g_1(Y_1, Y_2) = Y_1$ y $g(Y_1, Y_2) = -Y_2$, vemos que

$$E(Y_1 - Y_2) = E(Y_1) + E(-Y_2).$$

Se aplica el Teorema 5.7, dando $E(-Y_2) = -E(Y_2)$; por tanto,

$$E(Y_1 - Y_2) = E(Y_1) - E(Y_2).$$

También,

$$\begin{aligned} E(Y_1) &= \int_0^1 \int_0^{y_1} y_1(3y_1)\, dy_2\, dy_1 = \int_0^1 3y_1^3\, dy_1 = \frac{3}{4} y_1^4 \bigg]_0^1 = \frac{3}{4}, \\ E(Y_2) &= \int_0^1 \int_0^{y_1} y_2(3y_1)\, dy_2\, dy_1 = \int_0^1 3y_1 \left(\frac{y_2^2}{2} \bigg]_0^{y_1} \right) dy_1 = \int_0^1 \frac{3}{2} y_1^3\, dy_1 \\ &= \frac{3}{8} y_1^4 \bigg]_0^1 = \frac{3}{8}. \end{aligned}$$

Entonces,

$$E(Y_1 - Y_2) = (3/4) - (3/8) = 3/8,$$

de modo que esperaríamos que 3/8 del tanque esté lleno al final de las ventas de la semana. ∎

Si las variables aleatorias motivo de estudio son independientes, en ocasiones podemos simplificar el trabajo necesario para hallar valores esperados. El siguiente teorema es muy útil en este sentido.

TEOREMA 5.9 Sean Y_1 y Y_2 variables aleatorias independientes y sean $g(Y_1)$ y $h(Y_2)$ funciones sólo de Y_1 y Y_2, respectivamente. Entonces

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)],$$

siempre que existan los valores esperados.

Demostración Daremos la demostración del resultado para el caso continuo. Denotemos con $f(y_1, y_2)$ la densidad conjunta de Y_1 y Y_2. El producto $g(Y_1)h(Y_2)$ es una función de Y_1 y Y_2. Entonces, por la Definición 5.9 y la suposición de que Y_1 y Y_2 son independientes,

$$\begin{aligned}
E\left[g(Y_1)h(Y_2)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1)h(y_2)f(y_1, y_2)\,dy_2\,dy_1 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1)h(y_2)f_1(y_1)f_2(y_2)\,dy_2\,dy_1 \\
&= \int_{-\infty}^{\infty} g(y_1)f_1(y_1)\left[\int_{-\infty}^{\infty} h(y_2)f_2(y_2)\,dy_2\right]dy_1 \\
&= \int_{-\infty}^{\infty} g(y_1)f_1(y_1)E\left[h(Y_2)\right]dy_1 \\
&= E\left[h(Y_2)\right]\int_{-\infty}^{\infty} g(y_1)f_1(y_1)\,dy_1 = E\left[g(Y_1)\right]E\left[h(Y_2)\right].
\end{aligned}$$

La demostración para el caso discreto sigue un modo análogo.

EJEMPLO **5.21** Consulte el Ejemplo 5.19. En ese ejemplo encontramos $E(Y_1Y_2)$ directamente. Al investigar la forma de la función de densidad conjunta dada ahí, podemos ver que Y_1 y Y_2 son independientes. Encuentre $E(Y_1Y_2)$ con el uso del resultado de que $E(Y_1Y_2) = E(Y_1)\,E(Y_2)$ si Y_1 y Y_2 son independientes.

Solución La función de densidad conjunta está dada por

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Entonces,

$$f_1(y_1) = \begin{cases} \int_0^1 2(1 - y_1)\,dy_2 = 2(1 - y_1), & 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y

$$f_2(y_2) = \begin{cases} \int_0^1 2(1 - y_1)\,dy_1 = -(1 - y_1)^2\Big]_0^1 = 1, & 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Tenemos entonces

$$\begin{aligned}
E(Y_1) &= \int_0^1 y_1\left[2(1 - y_1)\right]dy_1 = 2\left(\frac{y_1^2}{2} - \frac{y_1^3}{3}\right)\Bigg]_0^1 = \frac{1}{3}, \\
E(Y_2) &= 1/2
\end{aligned}$$

porque Y_2 está uniformemente distribuida en (0, 1).

Se deduce que

$$E(Y_1Y_2) = E(Y_1)E(Y_2) = (1/3)(1/2) = 1/6,$$

lo que concuerda con la respuesta del Ejemplo 5.19. ■

Ejercicios

5.72 En el Ejercicio 5.1 determinamos que la distribución conjunta de Y_1, el número de contratos concedidos a la Empresa A, y Y_2, el número de contratos concedidos a la empresa B, está dada por las entradas de la tabla siguiente.

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

La función de probabilidad marginal de Y_1 se determinó en el Ejercicio 5.19 como la binomial con $n = 2$ y $p = 1/3$. Encuentre

a $E(Y_1)$.
b $V(Y_1)$.
c $E(Y_1 - Y_2)$.

5.73 En el Ejercicio 5.3 determinamos que la distribución de probabilidad conjunta de Y_1, el número de ejecutivos casados y Y_2, el número de ejecutivos que nunca se casaron, está dada por

$$p(y_1, y_2) = \frac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3 - y_1 - y_2}}{\binom{9}{3}},$$

donde y_1 y y_2 son enteros, $0 \le y_1 \le 3$, $0 \le y_2 \le 3$ y $1 \le y_1 + y_2 \le 3$. Encuentre el número esperado de ejecutivos casados de entre los tres seleccionados para promoción. (Vea el Ejercicio 5.21.)

5.74 Consulte los Ejercicios 5.6, 5.24 y 5.50. Suponga que una partícula radiactiva se coloca aleatoriamente en un cuadrado con lados de longitud unitaria. Un modelo razonable para la función de densidad conjunta para Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Cuál es $E(Y_1 - Y_2)$?
b ¿Cuál es $E(Y_1 Y_2)$?
c ¿Cuál es $E(Y_1^2 + Y_2^2)$?
d ¿Cuál es $V(Y_1 Y_2)$?

5.75 Consulte los Ejercicios 5.7, 5.25 y 5.51. Considere que Y_1 y Y_2 tienen la función de densidad conjunta

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & y_1 > 0, y_2 > 0 \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a ¿Cuáles son $E(Y_1 + Y_2)$ y $V(Y_1 + Y_2)$?
b ¿Cuál es $P(Y_1 - Y_2 > 3)$?
c ¿Cuál es $P(Y_1 - Y_2 < -3)$?
d ¿Cuáles son $E(Y_1 - Y_2)$ y $V(Y_1 - Y_2)$?
e ¿Qué se observa acerca de $V(Y_1 + Y_2)$ y $V(Y_1 - Y_2)$?

5.76 En el Ejercicio 5.8 dedujimos el hecho de que

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre $E(Y_1)$.
b Encuentre $V(Y_1)$.
c Encuentre $E(Y_1 - Y_2)$.

5.77 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \leq y_1 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. Encuentre

a $E(Y_1)$ y $E(Y_2)$.
b $V(Y_1)$ y $V(Y_2)$.
c $E(Y_1 - 3Y_2)$.

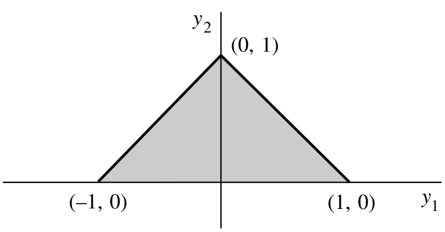
5.78 En el Ejercicio 5.10 demostramos que

$$f(y_1, y_2) = \begin{cases} 1, & 0 \leq y_1 \leq 2, 0 \leq y_2 \leq 1, 2y_2 \leq y_1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida para Y_1, la cantidad de contaminante por muestra recolectada arriba de la chimenea sin aparato limpiador y Y_2, la cantidad recolectada arriba de la chimenea con el limpiador.

a Encuentre $E(Y_1)$ y $E(Y_2)$.
b Encuentre $V(Y_1)$ y $V(Y_2)$.
c La variable aleatoria $Y_1 - Y_2$ representa la cantidad en la cual el peso de contaminante se puede reducir con el uso del aparato limpiador. Encuentre $E(Y_1 - Y_2)$.
d Encuentre $V(Y_1 - Y_2)$. ¿Dentro de qué límites espera que caiga $Y_1 - Y_2$?

5.79 Suponga que, como en el Ejercicio 5.11, Y_1 y Y_2 están uniformemente distribuidas sobre el triángulo sombreado en el diagrama siguiente. Encuentre $E(Y_1Y_2)$.



5.80 En el Ejercicio 5.16, Y_1 y Y_2 denotaron las proporciones del tiempo que los empleados I y II pasaban realmente trabajando en sus tareas asignadas durante un día hábil. La densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El empleado I tiene un porcentaje de productividad más alto que el II y una medida de la productividad total del par de empleados es $30Y_1 + 25Y_2$. Encuentre el valor esperado de esta medida de productividad.

5.81 En el Ejercicio 5.18, Y_1 y Y_2 denotaban la vida útil, en cientos de horas, para componentes de tipos I y II, respectivamente, en un sistema electrónico. La densidad conjunta de Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Una forma de medir la eficiencia relativa de los dos componentes es calcular la relación Y_2/Y_1. Encuentre $E(Y_2/Y_1)$. [*Sugerencia:* en el ejercicio 5.61 demostramos que Y_1 y Y_2 son independientes.]

5.82 En el Ejercicio 5.38 determinamos que la función de densidad conjunta para Y_1, el peso en toneladas de un artículo a granel que un proveedor tiene en existencia y Y_2, el peso del artículo vendido por el proveedor, tienen densidad conjunta

$$f(y_1, y_2) = \begin{cases} 1/y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En este caso la variable aleatoria $Y_1 - Y_2$ mide la cantidad del artículo remanente al final de la semana, una cantidad de gran importancia para el proveedor. Encuentre $E(Y_1 - Y_2)$.

5.83 En el Ejercicio 5.42 determinamos que la distribución de probabilidad incondicional para Y, el número de defectos por yarda en cierta tela, es

$$p(y) = (1/2)^{y+1}, \qquad y = 0, 1, 2, \ldots$$

Encuentre el número esperado de defectos por yarda.

5.84 En el Ejercicio 5.62 consideramos dos personas que lanzaban al aire una moneda hasta que aparecía la primera cara. Denote con Y_1 y Y_2 el número de veces que las personas A y B tiran la moneda, respectivamente. Si aparecen caras con probabilidad p y cruces con probabilidad $q = 1 - p$, es razonable concluir que Y_1 y Y_2 son independientes y que cada una tiene una distribución geométrica con parámetro p. Considere $Y_1 - Y_2$, la diferencia en el número de tiros que necesitan las dos personas.

a Encuentre $E(Y_1)$, $E(Y_2)$ y $E(Y_1 - Y_2)$.
b Encuentre $E(Y_1^2)$, $E(Y_2^2)$ y $E(Y_1Y_2)$ (recuerde que Y_1 y Y_2 son independientes).
c Encuentre $E(Y_1 - Y_2)^2$ y $V(Y_1 - Y_2)$.
d Proponga un intervalo que contenga $Y_1 - Y_2$ con probabilidad de al menos $8/9$.

5.85 En el Ejercicio 5.65 consideramos variables aleatorias Y_1 y Y_2 que, para $-1 \le \alpha \le 1$, tienen función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} [1 - \alpha\{(1 - 2e^{-y_1})(1 - 2e^{-y_2})\}]e^{-y_1-y_2}, & 0 \le y_1, 0 \le y_2, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

y establecimos que las distribuciones marginales de Y_1 y Y_2 eran exponenciales ambas con media 1. Encuentre

a $E(Y_1)$ y $E(Y_2)$.
b $V(Y_1)$ y $V(Y_2)$.

c $E(Y_1 - Y_2)$.

d $E(Y_1Y_2)$.

e $V(Y_1 - Y_2)$. ¿Dentro de qué límites se espera que caiga $Y_1 - Y_2$?

***5.86** Suponga que Z es una variable aleatoria normal estándar y que Y_1 y Y_2 son variables aleatorias con distribución χ^2 con ν_1 y ν_2 grados de libertad, respectivamente. Además, suponga que Z, Y_1 y Y_2 son independientes.

a Defina $W = Z/\sqrt{Y_1}$. Encuentre $E(W)$ y $V(W)$. ¿Qué suposiciones se necesitan acerca del valor de ν_1? [*Sugerencia:* $W = Z(1/\sqrt{Y_1}) = g(Z)h(Y_1)$. Use el Teorema 5.9. Los resultados del Ejercicio 4.112 d también serán útiles.]

b Defina $U = Y_1/Y_2$. Encuentre $E(U)$ y $V(U)$. ¿Qué suposiciones acerca de ν_1 y ν_2 se necesitan? Use la sugerencia del inciso a.

5.87 Suponga que Y_1 y Y_2 son variables aleatorias χ^2 independientes con ν_1 y ν_2 grados de libertad, respectivamente. Encuentre

a $E(Y_1 + Y_2)$.

b $V(Y_1 + Y_2)$. [*Sugerencia:* use el Teorema 5.9 y el resultado del Ejercicio 4.112a.]

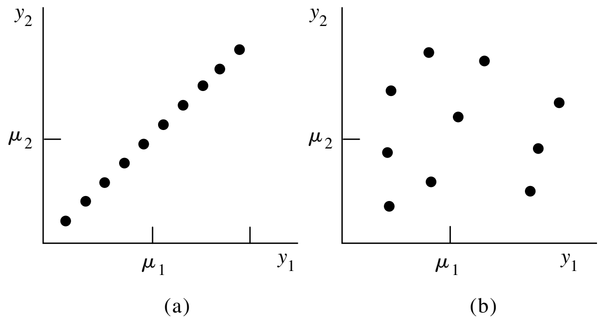
5.88 Supongamos que lanza un dado hasta que hayan salido cada una de las seis caras. ¿Cuál es el número esperado de tiros necesario para completar su tarea? [*Sugerencia:* si Y es el número de intentos para completar la tarea, $Y = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6$, donde Y_1 es el intento en el que cae la primera cara del dado, $Y_1 = 1$, Y_2 es el número de tiros adicionales necesario para que salga una cara diferente a la primera, Y_3 es el número de tiros adicionales necesario para que salga una cara diferente a las dos primeras caras distintas, . . . , Y_6 es el número de tiros adicionales necesario para que salga la última cara restante, después que todas las otras se hayan visto. Observe además que para $i = 2, \ldots, 6$, Y_i tiene una distribución geométrica con probabilidad de éxito $(7 - i)/6$.]

5.7 Covarianza de dos variables aleatorias

Intuitivamente consideramos la dependencia de dos variables aleatorias Y_1 y Y_2 como un proceso en el que una de las variables, por ejemplo Y_1, aumenta o disminuye cuando Y_2 cambia. Concentraremos nuestra atención en dos medidas de dependencia: la covarianza entre dos variables aleatorias y su coeficiente de correlación. En la Figura 5.8(a) y (b), se muestran las gráficas de los valores observados de dos variables, Y_1 y Y_2, para muestras de $n = 10$ unidades experimentales tomadas de cada una de las dos poblaciones. Si todos los puntos caen a lo largo de una recta, como indica la Figura 5.8(a), Y_1 y Y_2 son obviamente dependientes. En contraste, la Figura 5.8(b) indica poca o ninguna dependencia entre Y_1 y Y_2.

Suponga que conocemos los valores de $E(Y_1) = \mu_1$ y $E(Y_2) = \mu_2$ y localizamos este punto en la gráfica de la Figura 5.8. Ahora localizamos un punto graficado, (y_1, y_2), en la Figura 5.8(a) y medimos las desviaciones $(y_1 - \mu_1)$ y $(y_2 - \mu_2)$. Ambas desviaciones toman el mismo signo algebraico para cualquier punto, (y_1, y_2), y su producto $(y_1 - \mu_1)(y_2 - \mu_2)$ es positivo. Los puntos a la derecha de μ_1 generan pares de desviaciones positivas; los puntos a la izquierda producen pares de desviaciones negativas; y el promedio del producto de las desviaciones $(y_1 - \mu_1)(y_2 - \mu_2)$ es grande y positivo. Si la relación lineal indicada en la Figura 5.8(a) se hubiera inclinado hacia abajo a la derecha, todos los pares de desviaciones correspondientes hubieran sido de signo contrario y el valor promedio de $(y_1 - \mu_1)(y_2 - \mu_2)$ hubiera sido un número negativo grande.

FIGURA 5.8 Observaciones dependientes e independientes para (y_1, y_2)



La situación que acabamos de describir no ocurre para la Figura 5.8(b), donde existe poca dependencia entre Y_1 y Y_2. Sus desviaciones correspondientes $(y_1 - \mu_1)$ y $(y_2 - \mu_2)$ tomarán el mismo signo algebraico para algunos puntos y signos opuestos para otros. Entonces, el producto $(y_1 - \mu_1)(y_2 - \mu_2)$ será positivo para algunos puntos, negativo para otros y promediará algún valor cercano a cero.

Es evidente que el valor promedio de $(Y_1 - \mu_1)(Y_2 - \mu_2)$ proporciona una medida de la dependencia lineal entre Y_1 y Y_2. Esta cantidad, $E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$, se denomina *covarianza* de Y_1 y Y_2.

DEFINICIÓN **5.10**

Si Y_1 y Y_2 son variables aleatorias con medias μ_1 y μ_2, respectivamente, la *covarianza* de Y_1 y Y_2 es

$$\text{Cov}(Y_1, Y_2) = E\left[(Y_1 - \mu_1)(Y_2 - \mu_2)\right].$$

Cuanto mayor sea el valor absoluto de la covarianza de Y_1 y Y_2, mayor será la dependencia lineal entre Y_1 y Y_2. Los valores positivos indican que Y_1 aumenta cuando Y_2 aumenta; los valores negativos indican que Y_1 disminuye cuando Y_2 aumenta. Un valor cero de la covarianza indica que las variables son *no correlacionadas* y que no hay dependencia lineal entre Y_1 y Y_2.

Desafortunadamente, es difícil utilizar la covarianza como medida absoluta de dependencia porque su valor depende de la escala de medición. En consecuencia, es difícil determinar a primera vista si una covarianza particular es grande o pequeña. Este problema se puede eliminar al estandarizar su valor y usar el *coeficiente de correlación,* ρ, una cantidad relacionada con la varianza y que se define como

$$\rho = \frac{\text{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

donde σ_1 y σ_2 son desviaciones estándar de Y_1 y Y_2, respectivamente. Se pueden hallar más exposiciones del coeficiente de correlación en la obra de Hogg, Craig y McKean (2005) y Myers (2000).

Una demostración del coeficiente de correlación ρ satisface la desigualdad $-1 \leq \rho \leq 1$ está resumida en el Ejercicio 5.167.

El signo del coeficiente de correlación es igual al signo de la covarianza. Entonces, $\rho > 0$ indica que Y_2 aumenta a medida que Y_2 aumenta y $\rho = +1$ implica correlación perfecta, con todos los puntos cayendo en una recta con pendiente positiva. Un valor de $\rho = 0$ implica cero covarianza y que no hay correlación. Un coeficiente negativo de correlación implica una disminución en Y_2 cuando Y_1 aumenta, y $\rho = -1$ implica correlación perfecta, con todos los puntos cayendo en una recta con pendiente negativa. Una fórmula computacional conveniente para la covarianza se especifica en el siguiente teorema.

TEOREMA **5.10** Si Y_1 y Y_2 son variables aleatorias con medias μ_1 y μ_2, respectivamente, entonces

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] = E(Y_1 Y_2) - E(Y_1)E(Y_2).$$

Demostración

$$\begin{aligned}\text{Cov}(Y_1, Y_2) &= E[(Y_1 - \mu_1)(Y_2 - \mu_2)] \\ &= E(Y_1 Y_2 - \mu_1 Y_2 - \mu_2 Y_1 + \mu_1 \mu_2).\end{aligned}$$

Del Teorema 5.8, el valor esperado de una suma es igual a la suma de los valores esperados; y del Teorema 5.7, el valor esperado de una constante multiplicado por una función de variables aleatorias es la constante por el valor esperado. Entonces,

$$\text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - \mu_1 E(Y_2) - \mu_2 E(Y_1) + \mu_1 \mu_2.$$

Como $E(Y_1) = \mu_1$ y $E(Y_2) = \mu_2$, se deduce que

$$\text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = E(Y_1 Y_2) - \mu_1 \mu_2.$$

EJEMPLO **5.22** Consulte el Ejemplo 5.4. Encuentre la covarianza entre la cantidad en existencia Y_1 y la cantidad de ventas Y_2.

Solución Recuerde que Y_1 y Y_2 tienen función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Entonces,

$$\begin{aligned} E(Y_1 Y_2) &= \int_0^1 \int_0^{y_1} y_1 y_2 (3y_1)\, dy_2\, dy_1 = \int_0^1 3y_1^2 \left(\frac{y_2^2}{2} \bigg]_0^{y_1} \right) dy_1 \\ &= \int_0^1 \frac{3}{2} y_1^4\, dy_1 = \frac{3}{2} \left(\frac{y_1^5}{5} \bigg]_0^1 \right) = \frac{3}{10}. \end{aligned}$$

Del Ejemplo 5.20, sabemos que $E(Y_1) = 3/4$ y $E(Y_2) = 3/8$. Entonces, usando el Teorema 5.10, obtenemos

$$\text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = (3/10) - (3/4)(3/8) = .30 - .28 = .02.$$

En este ejemplo, valores grandes de Y_2 pueden presentarse sólo con valores grandes de Y_1 y la densidad, $f(y_1, y_2)$, es más grande para valores más grandes de Y_1 (vea la Figura 5.4). Entonces, intuimos que la covarianza entre Y_1 y Y_2 debe ser positiva. ■

EJEMPLO **5.23** Tengan Y_1 y Y_2 densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la covarianza de Y_1 y Y_2.

Solución Del Ejemplo 5.15, $E(Y_1Y_2) = 1/3$. También, de los Ejemplos 5.16 y 5.17, $\mu_1 = E(Y_1) = 2/3$ y $\mu_2 = E(Y_2) = 1/2$, y

$$\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - \mu_1\mu_2 = (1/3) - (2/3)(1/2) = 0.$$ ■

El Ejemplo 5.23 proporciona un caso específico del resultado general dado en el Teorema 5.11.

TEOREMA **5.11** Si Y_1 y Y_2 son variables aleatorias independientes, entonces

$$\text{Cov}(Y_1, Y_2) = 0.$$

Así, las variables aleatorias independientes deben ser no correlacionadas.

Demostración El Teorema 5.10 establece que

$$\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - \mu_1\mu_2.$$

Como Y_1 y Y_2 son independientes, el Teorema 5.9 implica que

$$E(Y_1Y_2) = E(Y_1)E(Y_2) = \mu_1\mu_2,$$

y el resultado deseado se deduce de inmediato.

Observe que las variables aleatorias Y_1 y Y_2 del Ejemplo 5.23 son independientes; en consecuencia, por el Teorema 5.11, su covarianza debe ser cero. El recíproco del Teorema 5.11 no es verdadero, como se ilustra en el ejemplo siguiente.

EJEMPLO **5.24** Sean Y_1 y Y_2 variables aleatorias discretas con distribución de probabilidad conjunta como se ve en la Tabla 5.3. Demuestre que Y_1 y Y_2 son dependientes pero tienen covarianza cero.

Solución El cálculo de probabilidades marginales da $p_1(-1) = p_1(1) = 5/16 = p_2(-1) = p_2(1)$ y $p_1(0) = 6/16 = p_2(0)$. El valor $p(0, 0) = 0$ en la celda del centro se destaca. Obviamente,

Tabla 5.3 Distribución de probabilidad conjunta, Ejemplo 5.24

	y_1		
y_2	-1	0	$+1$
-1	1/16	3/16	1/16
0	3/16	0	3/16
$+1$	1/16	3/16	1/16

$$p(0, 0) \neq p_1(0)p_2(0),$$

y esto es suficiente para demostrar que Y_1 y Y_2 son dependientes.

Observando de nuevo las probabilidades marginales, vemos que $E(Y_1) = E(Y_2) = 0$. También,

$$\begin{aligned} E(Y_1Y_2) &= \sum_{\text{toda } y_1} \sum_{\text{toda } y_2} y_1 y_2 p(y_1, y_2) \\ &= (-1)(-1)(1/16) + (-1)(0)(3/16) + (-1)(1)(1/16) \\ &\quad + (0)(-1)(3/16) + (0)(0)(0) + (0)(1)(3/16) \\ &\quad + (1)(-1)(1/16) + (1)(0)(3/16) + (1)(1)(1/16) \\ &= (1/16) - (1/16) - (1/16) + (1/16) = 0. \end{aligned}$$

Entonces,

$$\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = 0 - 0(0) = 0.$$

Este ejemplo demuestra que el recíproco del Teorema 5.11 no es verdadero. Si la covarianza de dos variables aleatorias es cero, las variables *no necesitan* ser independientes. ■

Ejercicios

5.89 En el Ejercicio 5.1 determinamos que la distribución conjunta de Y_1, el número de contratos concedidos a la empresa A, y Y_2, el número de contratos concedidos a la empresa B, está dada por las entradas en la tabla siguiente.

	y_1		
y_2	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

Encuentre $\text{Cov}(Y_1, Y_2)$. ¿Le sorprende que $\text{Cov}(Y_1, Y_2)$ sea negativa? ¿Por qué?

5.90 En el Ejercicio 5.3 determinamos que la distribución de probabilidad conjunta de Y_1, el número de ejecutivos casado, y Y_2, el número de ejecutivos nunca casados, está dada por

$$p(y_1, y_2) = \frac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3 - y_1 - y_2}}{\binom{9}{3}},$$

donde y_1 y y_2 son enteros, $0 \leq y_1 \leq 3$, $0 \leq y_2 \leq 3$ y $1 \leq y_1 + y_2 \leq 3$. Encuentre $\text{Cov}(Y_1, Y_2)$.

5.91 En el Ejercicio 5.8 dedujimos el hecho que

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

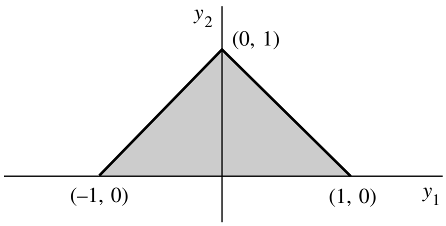
Demuestre que $\text{Cov}(Y_1, Y_2) = 0$. ¿Le sorprende que $\text{Cov}(Y_1, Y_2)$ sea cero? ¿Por qué?

5.92 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta valida. Encuentre Cov(Y_1, Y_2). ¿Y_1 y Y_2 son independientes?

5.93 Suponga que, al igual que en los Ejercicios 5.11 y 5.79, Y_1 y Y_2 están uniformemente distribuidas sobre el triángulo sombreado del diagrama siguiente.



a Encuentre Cov(Y_1, Y_2).
b ¿Y_1 y Y_2 son independientes? (Vea el Ejercicio 5.55.)
c Encuentre el coeficiente de correlación para Y_1 y Y_2.
d ¿La respuesta al inciso b lo lleva a dudar de su respuesta al inciso a? ¿Por qué sí o por qué no?

5.94 Sean Y_1 y Y_2 variables aleatorias no correlacionadas y considere $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$.

a Encuentre la Cov(U_1, U_2) en términos de las varianzas de Y_1 y Y_2.
b Encuentre una expresión para el coeficiente de correlación entre U_1 y U_2.
c ¿Es posible que Cov(U_1, U_2) = 0? ¿Cuándo ocurre esto?

5.95 Consideremos que las variables aleatorias discretas Y_1 y Y_2 tienen la función de probabilidad conjunta

$$p(y_1, y_2) = 1/3, \qquad \text{para } (y_1, y_2) = (-1, 0), (0, 1), (1, 0).$$

Encuentre Cov(Y_1, Y_2). Observe que Y_1 y Y_2 son dependientes. (¿Por qué?) Éste es otro ejemplo de variables aleatorias no correlacionadas que no son independientes.

5.96 Suponga que las variables aleatorias Y_1 y Y_2 tienen medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Use la definición básica de la covarianza de dos variables aleatorias para establecer que

a $\text{Cov}(Y_1,\ Y_2) = \text{Cov}(Y_2,\ Y_1)$.
b $\text{Cov}(Y_1,\ Y_1) = V(Y_1) = \sigma_1^2$. Esto es, la covarianza de una variable aleatoria y ella misma son sólo la varianza de la variable aleatoria.

5.97 Las variables aleatorias Y_1 y Y_2 son tales que $E(Y_1) = 4$, $E(Y_2) = -1$, $V(Y_1) = 2$ y $V(Y_2) = 8$.

a ¿Cuál es Cov(Y_1, Y_1)?
b Suponiendo que las medias y las varianzas sean correctas, ¿es posible que Cov(Y_1, Y_2) = 7? [*Sugerencia:* si Cov(Y_1, Y_2) = 7, ¿cuál es el valor de ρ, el coeficiente de correlación?]
c Suponiendo que las medias y las varianzas sean correctas, ¿cuál es el máximo valor posible para Cov(Y_1, Y_2)? Si Cov(Y_1, Y_2) alcanza este valor máximo, ¿qué implica eso acerca de la relación entre Y_1 y Y_2?

d Suponiendo que las medias y las varianzas sean correctas, ¿cuál es el mínimo valor posible para $\text{Cov}(Y_1, Y_2)$? Si $\text{Cov}(Y_1, Y_2)$ alcanza este valor mínimo, ¿qué implica esto acerca de la relación entre Y_1 y Y_2?

5.98 ¿Qué tan grande o pequeña puede ser $\text{Cov}(Y_1, Y_2)$? Use el dato de que $\rho^2 \le 1$ para demostrar que

$$-\sqrt{V(Y_1) \times V(Y_2)} \le \text{Cov}(Y_1, Y_2) \le \sqrt{V(Y_1) \times V(Y_2)}.$$

5.99 Si c es cualquier constante y Y es una variable aleatoria tal que $E(Y)$ existe, demuestre que $\text{Cov}(c, Y) = 0$.

5.100 Sea Z una variable aleatoria normal estándar y sean $Y_1 = Z$ y $Y_2 = Z^2$.

a ¿Cuáles son $E(Y_1)$ y $E(Y_2)$?

b ¿Cuál es $E(Y_1Y_2)$? [*Sugerencia:* $E(Y_1Y_2) = E(Z^3)$, recuerde el Ejercicio 4.199.]

c ¿Cuál es $\text{Cov}(Y_1, Y_2)$?

d Observe que $P(Y_2 > 1 | Y_1 > 1) = 1$. ¿Y_1 y Y_2 son independientes?

5.101 En el Ejercicio 5.65 consideramos variables aleatorias Y_1 y Y_2 que, para $-1 \le \alpha \le 1$, tienen función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} [1 - \alpha\{(1 - 2e^{-y_1})(1 - 2e^{-y_2})\}]e^{-y_1 - y_2}, & 0 \le y_1, 0 \le y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Establecimos que las distribuciones marginales de Y_1 y Y_2 son ambas exponenciales con media 1, y demostramos que Y_1 y Y_2 son independientes si y sólo si $\alpha = 0$. En el Ejercicio 5.85 dedujimos $E(Y_1Y_2)$.

a Derive $\text{Cov}(Y_1, Y_2)$.

b Demuestre que $\text{Cov}(Y_1, Y_2) = 0$ si y sólo si $\alpha = 0$.

c Demuestre que Y_1 y Y_2 son independientes si y sólo si $\rho = 0$.

5.8 Valor esperado y varianza de funciones lineales de variables aleatorias

Más adelante en este texto, en especial los Capítulos 9 y 11, frecuentemente encontraremos estimadores que son funciones lineales de las mediciones en una muestra, $Y_1, Y_2, \ldots, Y_n$. Si $a_1, a_2 \ldots, a_n$ son constantes, será necesario calcular el valor esperado y varianza de una función lineal de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$.

$$U_1 = a_1Y_1 + a_2Y_2 + a_3Y_3 + \ldots + a_nY_n = \sum_{i=1}^{n} a_iY_i.$$

También podemos estar interesados en la covarianza entre dos de estas combinaciones lineales. Los resultados que simplifican el cálculo de estas cantidades se resumen en el teorema siguiente.

TEOREMA 5.12 Sean $Y_1, Y_2, \ldots, Y_n$ y $X_1, X_2, \ldots, X_m$ variables aleatorias con $E(Y_i) = \mu_i$ y $E(X_j) = \xi_j$. Defina

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{y} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

para las constantes $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots, b_m$. Entonces se cumple lo siguiente:

a $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.

b $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{1 \le i < j \le n} a_i a_j \text{Cov}(Y_i, Y_j)$, donde la doble suma es para todos los pares (i, j) con $i < j$.

c $\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$.

Antes de continuar con la demostración del Teorema 5.12, ejemplificaremos su uso.

EJEMPLO 5.25 Sean Y_1, Y_2 y Y_3 variables aleatorias, donde $E(Y_1) = 1$, $E(Y_2) = 2$, $E(Y_3) = -1$, $V(Y_1) = 1$, $V(Y_2) = 3$, $V(Y_3) = 5$, $\text{Cov}(Y_1, Y_2) = -0.4$, $\text{Cov}(Y_1, Y_3) = 1/2$ y $\text{Cov}(Y_2, Y_3) = 2$. Encuentre el valor esperado y la varianza de $U = Y_1 - 2Y_2 + Y_3$. Si $W = 3Y_1 + Y_2$, encuentre $\text{Cov}(U, W)$.

Solución $U = a_1Y_1 + a_2Y_2 + a_3Y_3$, donde $a_1 = 1$, $a_2 = -2$ y $a_3 = 1$. Entonces, por el Teorema 5.12,

$$E(U) = a_1 E(Y_1) + a_2 E(Y_2) + a_3 E(Y_3) = (1)(1) + (-2)(2) + (1)(-1) = -4.$$

De manera similar,

$$\begin{aligned} V(U) &= a_1^2 V(Y_1) + a_2^2 V(Y_2) + a_3^2 V(Y_3) + 2a_1a_2\text{Cov}(Y_1, Y_2) \\ &\quad + 2a_1a_3\text{Cov}(Y_1, Y_3) + 2a_2a_3\text{Cov}(Y_2, Y_3) \\ &= (1)^2(1) + (-2)^2(3) + (1)^2(5) + (2)(1)(-2)(-0.4) \\ &\quad + (2)(1)(1)(1/2) + (2)(-2)(1)(2) \\ &= 12.6. \end{aligned}$$

Observe que $W = b_1Y_1 + b_2Y_2$, donde $b_1 = 3$ y $b_2 = 1$. Entonces,

$$\begin{aligned} \text{Cov}(U, W) &= a_1b_1\text{Cov}(Y_1, Y_1) + a_1b_2\text{Cov}(Y_1, Y_2) + a_2b_1\text{Cov}(Y_2, Y_1) \\ &\quad + a_2b_2\text{Cov}(Y_2, Y_2) + a_3b_1\text{Cov}(Y_3, Y_1) + a_3b_2\,\text{Cov}(Y_3, Y_2). \end{aligned}$$

Observe que, como se establece en el Ejercicio 5.96, $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$ y $\text{Cov}(Y_i, Y_i) = V(Y_i)$. Por tanto,

$$\begin{aligned}\text{Cov}(U, W) &= (1)(3)(1) + (1)(1)(-0.4) + (-2)(3)(-0.4)\\ &\quad + (-2)(1)(3) + (1)(3)(1/2) + (1)(1)(2)\\ &= 2.5.\end{aligned}$$

Como $\text{Cov}(U, W) \neq 0$, se deduce que U y W son *dependientes*. ■

Ahora continuamos con la demostración del Teorema 5.12.

Demostración

El teorema consta de tres partes, de las cuales (a) viene directamente de los Teoremas 5.7 y 5.8. Para demostrar (b) recurrimos a la definición de varianza y escribimos

$$\begin{aligned}V(U_1) = E\,[U_1 - E(U_1)]^2 &= E\left[\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right]^2\\ &= E\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]^2\\ &= E\left[\sum_{i=1}^{n} a_i^2 (Y_i - \mu_i)^2 + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_i a_j (Y_i - \mu_i)(Y_j - \mu_j)\right]\\ &= \sum_{i=1}^{n} a_i^2 E(Y_i - \mu_i)^2 + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_i a_j E\left[(Y_i - \mu_i)(Y_j - \mu_j)\right].\end{aligned}$$

Por las definiciones de varianza y covarianza tenemos

$$V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + \sum_{i=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} a_i a_j \,\text{Cov}(Y_i, Y_j).$$

Como $\text{Cov}(Y_i, Y_j) = \text{Cov}(Y_j, Y_i)$, podemos escribir

$$V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{1\leq i<j\leq n} a_i a_j \,\text{Cov}(Y_i, Y_j).$$

Se pueden usar pasos similares para obtener (c). Tenemos

$$\begin{aligned}\text{Cov}(U_1, U_2) &= E\{[U_1 - E(U_1)]\,[U_2 - E(U_2)]\}\\ &= E\left[\left(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{m} b_j X_j - \sum_{j=1}^{m} b_j \xi_j\right)\right]\\ &= E\left\{\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]\left[\sum_{j=1}^{m} b_j (X_j - \xi_j)\right]\right\}\end{aligned}$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j (Y_i - \mu_i)(X_j - \xi_j)\right]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(Y_i - \mu_i)(X_j - \xi_j)]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \operatorname{Cov}(Y_i, X_j).$$

Al observar que $\operatorname{Cov}(Y_i, Y_i) = V(Y_i)$, podemos ver que (b) es un caso especial de (c).

EJEMPLO **5.26** Consulte los Ejemplos 5.4 y 5.20. En el segundo estuvimos interesados en $Y_1 - Y_2$, la cantidad proporcional de gasolina restante al final de una semana. Encuentre la varianza de $Y_1 - Y_2$.

Solución Usando el Teorema 5.12, tenemos

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2\operatorname{Cov}(Y_1, Y_2).$$

Como

$$f_1(y_1) = \begin{cases} 3y_1^2, & 0 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y

$$f_2(y_2) = \begin{cases} (3/2)(1 - y_2^2), & 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

se deduce que

$$E(Y_1^2) = \int_0^1 3y_1^4\, dy_1 = \frac{3}{5},$$
$$E(Y_2^2) = \int_0^1 \frac{3}{2} y_2^2(1 - y_2^2)\, dy_2 = \frac{3}{2}\left[\frac{1}{3} - \frac{1}{5}\right] = \frac{1}{5}.$$

Del Ejemplo 5.20 tenemos $E(Y_1) = 3/4$ y $E(Y_2) = 3/8$. Entonces,

$$V(Y_1) = (3/5) - (3/4)^2 = .04 \quad \text{y} \quad V(Y_2) = (1/5) - (3/8)^2 = .06.$$

En el Ejemplo 5.22 determinamos que $\operatorname{Cov}(Y_1, Y_2) = .02$. Por tanto,

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2\operatorname{Cov}(Y_1, Y_2)$$
$$= .04 + .06 - 2(.02) = .06.$$

La desviación estándar de $Y_1 - Y_2$ es entonces $\sqrt{.06} = .245$. ■

EJEMPLO **5.27** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$. (Estas variables pueden denotar los resultados de n intentos independientes de un experimento.) Defina

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

y demuestre que $E(\overline{Y}) = \mu$ y $V(\overline{Y}) = \sigma^2/n$.

Solución Observe que $\overline{Y}$ es una función lineal de $Y_1, Y_2, \ldots, Y_n$ con todas las constantes a_i iguales a $1/n$. Esto es,

$$\overline{Y} = \left(\frac{1}{n}\right) Y_1 + \cdots + \left(\frac{1}{n}\right) Y_n.$$

Por el Teorema 5.12(a),

$$E(\overline{Y}) = \sum_{i=1}^{n} a_i\mu_i = \sum_{i=1}^{n} a_i\mu = \mu\sum_{i=1}^{n} a_i = \mu\sum_{i=1}^{n}\frac{1}{n} = \frac{n\mu}{n} = \mu.$$

Por el Teorema 5.12(b),

$$V(\overline{Y}) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2\sum_{\substack{i=1\\ i<j}}^{n}\sum_{i=1}^{n} a_i a_j \operatorname{Cov}(Y_i, Y_j).$$

Los términos de la covarianza son todos iguales a cero porque las variables aleatorias son independientes. Entonces,

$$V(\overline{Y}) = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2\sigma_i^2 = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2\sigma^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$ ■

EJEMPLO **5.28** El número de artículos defectuosos Y en una muestra de $n = 10$ artículos seleccionados del proceso de fabricación tiene una distribución de probabilidad binomial. Un estimador de la fracción defectuosa del lote es la variable $\hat{p} = Y/n$. Encuentre el valor esperado y la varianza de $\hat{p}$.

Solución El término $\hat{p}$ es una función lineal de una sola variable Y, donde $\hat{p} = a_1 Y$ y $a_1 = 1/n$. Entonces, por el Teorema 5.12,

$$E(\hat{p}) = a_1 E(Y) = \frac{1}{n}E(Y).$$

El valor esperado y la varianza de una variable aleatoria binomial son np y npq, respectivamente. Sustituyendo por $E(Y)$, obtenemos

$$E(\hat{p}) = \frac{1}{n}(np) = p.$$

Entonces, el valor esperado del número de artículos defectuosos Y, dividido entre el tamaño muestral, es p. Del mismo modo

$$V(\hat{p}) = a_1^2 V(Y) = \left(\frac{1}{n}\right)^2 npq = \frac{pq}{n}. \qquad \blacksquare$$

EJEMPLO **5.29** Suponga que una urna contiene r bolas rojas y $(N - r)$ bolas negras. Una muestra aleatoria de n bolas se saca sin restitución y se observa Y, el número de bolas rojas de la muestra. Del Capítulo 3 sabemos que Y tiene una distribución de probabilidad hipergeométrica. Encuentre la media y la varianza de Y.

Solución Primero observamos algunas características de muestrear sin restitución. Suponga que el muestreo se realiza en forma secuencial y observamos resultados para $X_1, X_2, \dots, X_n$, donde

$$X_i = \begin{cases} 1, & \text{si el } i\text{-ésimo resultado es una bola roja,} \\ 0, & \text{en caso contrario.} \end{cases}$$

Incuestionablemente, $P(X_1 = 1) = r/N$. Pero también es cierto que $P(X_2 = 1) = r/N$ porque

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1, X_2 = 1) + P(X_1 = 0, X_2 = 1) \\ &= P(X_1 = 1)P(X_2 = 1|X_1 = 1) + P(X_1 = 0)P(X_2 = 1|X_1 = 0) \\ &= \left(\frac{r}{N}\right)\left(\frac{r-1}{N-1}\right) + \left(\frac{N-r}{N}\right)\left(\frac{r}{N-1}\right) = \frac{r(N-1)}{N(N-1)} = \frac{r}{N}. \end{aligned}$$

Lo mismo es cierto para X_k; esto es,

$$P(X_k = 1) = \frac{r}{N}, \qquad k = 1, 2, \dots, n.$$

Entonces, la probabilidad (incondicional) de sacar una bola roja en cualquier ensayo es r/N.

Del mismo modo, se puede demostrar que

$$P(X_j = 1, X_k = 1) = \frac{r(r-1)}{N(N-1)}, \qquad j \neq k.$$

Ahora, observe que $Y = \sum_{i=1}^{n} X_i$ y por lo tanto,

$$E(Y) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \left(\frac{r}{N}\right) = n\left(\frac{r}{N}\right).$$

Para hallar $V(Y)$ necesitamos $V(X_i)$ y $\text{Cov}(X_i, X_j)$. Como X_i es 1 con probabilidad r/N y 0 con probabilidad $1 - (r/N)$, se deduce que

$$V(X_i) = \frac{r}{N}\left(1 - \frac{r}{N}\right).$$

También,

$$\operatorname{Cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)^2$$
$$= -\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{1}{N-1}\right)$$

porque $X_i X_j = 1$ si y sólo si $X_i = 1$ y $X_j = 1$ y $X_1 X_j = 0$ en caso contrario. Del Teorema 5.12, sabemos que

$$V(Y) = \sum_{i=1}^{n} V(X_i) + 2\sum\sum_{i<j} \operatorname{Cov}(X_i, X_j)$$
$$= \sum_{i=1}^{n}\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right) + 2\sum\sum_{i<j}\left[-\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{1}{N-1}\right)\right]$$
$$= n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right) - n(n-1)\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{1}{N-1}\right)$$

porque la doble sumatoria contiene $n(n-1)/2$ términos iguales. Un poco de álgebra dará como resultado

$$V(Y) = n\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

■

Para apreciar la utilidad del Teorema 5.12, Observe que las deducciones contenidas en el Ejemplo 5.29 son mucho más sencillas que las resumidas en el Ejercicio 3.216, donde la media y la varianza se dedujeron usando las probabilidades asociadas con la distribución hipergeométrica.

Ejercicios

5.102 Una empresa compra dos tipos de productos químicos industriales. El producto tipo I cuesta \$3 el galón, mientras que el tipo II cuesta \$5 por galón. La media y la varianza para el número de galones comprado del producto tipo I, Y_1, son 40 y 4, respectivamente. La cantidad comprada del producto tipo II, Y_2, tiene $E(Y_2) = 65$ galones y $V(Y_2) = 8$. Suponga que Y_1 y Y_2 son independientes y encuentre la media y la varianza de la cantidad total de dinero gastado por semana en los dos productos químicos.

5.103 Suponga que Y_1, Y_2 y Y_3 son variables aleatorias, con

$$\begin{array}{rrr} E(Y_1) = 2, & E(Y_2) = -1, & E(Y_3) = 4, \\ V(Y_1) = 4, & V(Y_2) = 6, & V(Y_3) = 8, \\ \operatorname{Cov}(Y_1, Y_2) = 1, & \operatorname{Cov}(Y_1, Y_3) = -1, & \operatorname{Cov}(Y_2, Y_3) = 0. \end{array}$$

Encuentre $E(3Y_1 + 4Y_2 - 6Y_3)$ y $V(3Y_1 + 4Y_2 - 6Y_3)$.

5.104 En el Ejercicio 5.3 determinamos que la distribución de probabilidad conjunta de Y_1, el número de ejecutivos casados y Y_2, el número de ejecutivos nunca casados, está dada por

$$p(y_1, y_2) = \frac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3-y_1-y_2}}{\binom{9}{3}}$$

donde y_1 y y_2 son enteros, $0 \le y_1 \le 3$, $0 \le y_2 \le 3$ y $1 \le y_1 + y_2 \le 3$.

a Encuentre $E(Y_1 + Y_2)$ y $V(Y_1 + Y_2)$ hallando primero la distribución de probabilidad de $Y_1 + Y_2$.

b En el Ejercicio 5.90 determinamos que $\text{Cov}(Y_1, Y_2) = -1/3$. Encuentre $E(Y_1 + Y_2)$ y $V(Y_1 + Y_2)$ usando el Teorema 5.12.

5.105 En el Ejercicio 5.8 establecimos que

$$f(y_1, y_2) = \begin{cases} 4y_1y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. En el Ejercicio 5.52 establecimos que Y_1 y Y_2 son independientes; en el Ejercicio 5.76 determinamos que $E(Y_1 - Y_2) = 0$ y encontramos el valor para $V(Y_1)$. Encuentre $V(Y_1 - Y_2)$.

5.106 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \le y_1 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida. En el Ejercicio 5.76 dedujimos el hecho que $E(Y_1 - 3Y_2) = -5/4$; en el Ejercicio 5.92 demostramos que $\text{Cov}(Y_1, Y_2) = 1/40$. Encuentre $V(Y_1 - 3Y_2)$.

5.107 En el Ejercicio 5.12 se estableció la siguiente función de densidad de probabilidad conjunta para las variables aleatorias Y_1 y Y_2, que fueron las proporciones de dos componentes de una muestra de una mezcla de insecticida:

$$f(y_1, y_2) = \begin{cases} 2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, 0 \le y_1 + y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Para los dos productos químicos en consideración, una cantidad importante es la proporción total $Y_1 + Y_2$ hallada en cualquier muestra. Encuentre $E(Y_1 + Y_2)$ y $V(Y_1 + Y_2)$.

5.108 Si Y_1 es el tiempo total entre la llegada de un cliente a la tienda y su salida de la ventanilla de servicio, y si Y_2 es el tiempo que pasa en la fila de espera antes de llegar a la ventanilla, la densidad conjunta de estas variables se dio en el Ejercicio 5.15 como

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \le y_2 \le y_1 \le \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La variable aleatoria $Y_1 - Y_2$ representa el tiempo que pasa en la ventanilla de servicio. Encuentre $E(Y_1 - Y_2)$ y $V(Y_1 - Y_2)$. ¿Es altamente probable que un cliente seleccionado aleatoriamente pase más de 4 minutos en la ventanilla de servicio?

5.109 En el Ejercicio 5.16, Y_1 y Y_2 denotaron las proporciones de tiempo que los empleados I y II en realidad dedicaron a trabajar en sus tareas asignadas durante una jornada. La densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} y_1 + y_2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el Ejercicio 5.80 dedujimos la media de la medida de productividad $30Y_1 + 25Y_2$. Encuentre la varianza de esta medida de productividad. Proporcione un intervalo en el que piense que las medidas de productividad total de los dos empleados deben estar al menos 75% de los días en cuestión.

5.110 Suponga que Y_1 y Y_2 tienen coeficiente de correlación $\rho = .2$. ¿Cuál es el valor del coeficiente de correlación entre

a $1 + 2Y_1$ y $3 + 4Y_2$?
b $1 + 2Y_1$ y $3 - 4Y_2$?
c $1 - 2Y_1$ y $3 - 4Y_2$?

5.111 Suponga que Y_1 y Y_2 tienen coeficiente de correlación ρ_{Y_1, Y_2} y para las constantes a, b, c y d sea $W_1 = a + bY_1$ y $W_2 = c + dY_2$.

a Demuestre que el coeficiente de correlación entre W_1 y W_2, ρ_{W_1, W_2}, es tal que $|\rho_{Y_1, Y_2}| = |\rho_{W_1, W_2}|$.
b ¿Este resultado explica los resultados obtenidos en el Ejercicio 5.110?

5.112 En el Ejercicio 5.18, Y_1 y Y_2 denotaban las vidas útiles, en cientos de horas, para componentes de tipos I y II, respectivamente, en un sistema electrónico. La densidad conjunta de Y_1 y Y_2 es

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0, y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El costo C de cambiar los dos componentes depende de la duración de sus vidas útiles cuando fallan y está dada por $C = 50 + 2Y_1 + 4Y_2$. Encuentre $E(C)$ y $V(C)$.

5.113 Una comerciante minorista piensa que su ganancia diaria X, obtenida de sus ventas, es una variable aleatoria normalmente distribuida con $\mu = 50$ y $\sigma = 3$ (mediciones en dólares). X puede ser negativa si ella se ve forzada a deshacerse de suficientes artículos perecederos. Del mismo modo, calcula que los costos generales diarios Y tienen una distribución gamma con $\alpha = 4$ y $\beta = 2$. Si X y Y son independientes, encuentre el valor esperado y la varianza de la *ganancia* diaria neta de la comerciante. ¿Es de esperarse que su ganancia neta para mañana rebase los \$70?

5.114 Para la producción diaria de una operación industrial, denote con Y_1 la cantidad de ventas y Y_2, los costos, en miles de dólares. Suponga que las funciones de densidad para Y_1 y Y_2 están dadas por

$$f_1(y_1) = \begin{cases} (1/6)y_1^3 e^{-y_1}, & y_1 > 0, \\ 0, & y_1 \le 0, \end{cases} \quad \text{y} \quad f_2(y_2) = \begin{cases} (1/2)e^{-y_2/2}, & y_2 > 0, \\ 0, & y_2 \le 0. \end{cases}$$

La utilidad diaria está dada por $U = Y_1 - Y_2$.

a Encuentre $E(U)$.
b Suponiendo que Y_1 y Y_2 son independientes, encuentre $V(U)$.
c ¿Se esperaría que la utilidad diaria cayera por debajo de cero con mucha frecuencia? ¿Por qué?

5.115 Consulte el Ejercicio 5.88. Si Y denota el número de tiros del dado hasta que se vea cada una de las seis caras, $Y = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6$ donde Y_1 es el intento en el que se lanza la primera cara del dado, $Y_1 = 1$, Y_2 es el número de tiros adicionales necesario para que salga una cara diferente a la primera, Y_3 es el número de tiros adicionales necesario para que salga una cara diferente a las dos primeras caras distintas, . . . , Y_6 es el número de tiros adicionales necesario para que salga la última cara restante, después de que todas las otras se hayan visto.

a Demuestre que $\text{Cov}(Y_i, Y_j) = 0$, $i, j = 1, 2, \ldots, 6$, $i \neq j$.
b Use el Teorema 5.12 para hallar $V(Y)$.
c Proponga un intervalo que contenga a Y con probabilidad de al menos $3/4$.

5.116 Consulte el Ejercicio 5.75. Use el Teorema 5.12 para explicar por qué $V(Y_1 + Y_2) = V(Y_1 - Y_2)$.

***5.117** Una población de N lagartos se ha de muestrear para obtener una medida aproximada de la diferencia entre las proporciones de machos y hembras sexualmente maduros. Es evidente que este parámetro tiene importantes implicaciones para el futuro de la población. Suponga que n animales se han de muestrear sin restitución. Denote con Y_1 el número de hembras maduras y con Y_2 el de machos maduros de la muestra. Si la población contiene proporciones p_1 y p_2 de hembras y machos maduros, respectivamente (con $p_1 + p_2 < 1$), encuentre expresiones para

$$E\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right) \quad \text{y} \quad V\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right).$$

5.118 La carga total sostenida en la cimentación de concreto de un edificio proyectado es la suma de la carga muerta más la carga de ocupación. Suponga que la carga muerta X_1 tiene una distribución gamma con $\alpha_1 = 50$ y $\beta_1 = 2$, mientras que la carga de ocupación X_2 tiene una distribución gamma con $\alpha_2 = 20$ y $\beta_2 = 2$. (Las unidades son en miles de libras.) Suponga que X_1 y X_2 son independientes.

a Encuentre la media y la varianza de la carga sostenida total en la cimentación.

b Encuentre un valor para la carga sostenida que será rebasada con probabilidad menor que $1/16$.

5.9 Distribución de probabilidad multinomial

Recuerde del Capítulo 3 que una variable aleatoria binomial resulta de un experimento que consiste en n intentos con dos posibles resultados por intento. Con frecuencia encontramos situaciones similares en las que el número de posibles resultados por intento es más que dos. Por ejemplo, experimentos que comprenden tipos de sangre por lo general tienen al menos cuatro posibles resultados por intento. Experimentos que comprenden el muestreo de artículos defectuosos pueden clasificar el tipo de defectos observados en más de dos clases.

Un experimento multinomial es una generalización del experimento binomial.

DEFINICIÓN **5.11**

Un *experimento multinomial* posee las siguientes propiedades:

1. El experimento consta de n intentos idénticos.
2. El resultado de cada intento cae en una de k clases o celdas.
3. La probabilidad de que el resultado de un solo intento caiga en la celda i, es p_i, $i = 1, 2, \ldots, k$ y sigue siendo el mismo de un intento a otro. Observe que $p_1 + p_2 + p_3 + \cdots + p_k = 1$.
4. Los intentos son independientes.
5. Las variables aleatorias de interés son $Y_1, Y_2, \ldots, Y_k$, donde Y_i es igual al número de intentos para los cuales el resultado cae en la celda i. Observe que $Y_1 + Y_2 + Y_3 + \cdots + Y_k = n$.

La función de probabilidad conjunta para $Y_1, Y_2, \ldots, Y_k$ está dada por

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1!y_2!\cdots y_k!} p_1^{y_1} p_2^{y_2} \ldots p_k^{y_k},$$

donde

$$\sum_{i=1}^{k} p_i = 1 \quad \text{y} \quad \sum_{i=1}^{k} y_i = n.$$

Calcular la probabilidad de que los n intentos en un experimento multinomial resulten en $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k)$ es una excelente aplicación de los métodos probabilísticos del Capítulo 2. Dejamos este problema como ejercicio.

DEFINICIÓN 5.12

Suponga que $p_1, p_2, \ldots, p_k$ son tales que $\sum_{i=1}^{k} p_i = 1$, y $p_i > 0$ para $i = 1, 2, \ldots, k$. Se dice que las variables aleatorias $Y_1, Y_2, \ldots, Y_k$ tienen una *distribución multinomial* con parámetros n y $p_1, p_2, \ldots, p_k$ si la función de probabilidad conjunta de $Y_1, Y_2, \ldots, Y_k$ está dada por

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k},$$

donde, para cada i, $y_i = 0, 1, 2, \ldots, n$ y $\sum_{i=1}^{k} y_i = n$.

Muchos experimentos en los que aparece una clasificación son multinomiales. Por ejemplo, clasificar personas en cinco grupos resulta en una enumeración o cantidad correspondiente a cada una de cinco clases de ingreso. O bien, podríamos estar interesados en estudiar la reacción de ratones a un estímulo particular en un experimento psicológico. Si los ratones pueden reaccionar en una de tres formas cuando se aplica el estímulo, el experimento da el número de ratones que caigan en cada clase de reacción. Del mismo modo, un estudio de tránsito podría requerir un conteo y clasificación de los tipos de vehículos de motor que usan una sección de carretera. Un proceso industrial podría manufacturar artículos que caigan en una de tres clases de calidad: aceptables, de segunda y rechazos. Un estudiante de arte podría clasificar pinturas en una de k categorías de acuerdo con el estilo y periodo, o podríamos clasificar ideas filosóficas de los autores en un estudio de literatura. El resultado de una campaña publicitaria podría dar datos que indiquen una clasificación de reacciones del consumidor. Muchas observaciones de ciencias físicas no son sensibles a mediciones en una escala continua y, por tanto, resultan en datos enumerativos que corresponden a los números de observaciones que caen en varias clases.

Observe que el experimento binomial es un caso especial del experimento multinomial (cuando hay $k = 2$ clases).

EJEMPLO 5.30 De acuerdo con cifras de un censo reciente, las proporciones de adultos (personas de más de 18 años de edad) en Estados Unidos, asociados con cinco categorías de edades se dan en la siguiente tabla.

Edad	Proporción
18–24	.18
25–34	.23
35–44	.16
45–64	.27
65↑	.16

Si estas cifras son precisas y cinco adultos se muestrean aleatoriamente, encuentre la probabilidad de que la muestra contenga una persona entre las edades de 18 y 24, dos entre 25 y 34, y dos entre 45 y 64.

Solución Vamos a numerar las cinco clases de edad 1, 2, 3, 4 y 5 de arriba abajo y supondremos que las proporciones dadas son las probabilidades asociadas con cada una de las clases. Entonces deseamos hallar

$$p(y_1, y_2, y_3, y_4, y_5) = \frac{n!}{y_1!\, y_2!\, y_3!\, y_4!\, y_5!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} p_5^{y_5},$$

para $n = 5$ y $y_1 = 1$, $y_2 = 2$, $y_3 = 0$, $y_4 = 2$ y $y_5 = 0$. Sustituyendo estos valores en la fórmula para la función de probabilidad conjunta, obtenemos

$$\begin{aligned} p(1,\ 2,\ 0,\ 2,\ 0) &= \frac{5!}{1!\,2!\,0!\,2!\,0!}(.18)^1(.23)^2(.16)^0(.27)^2(.16)^0 \\ &= 30(.18)(.23)^2(.27)^2 = .0208. \end{aligned}$$

■

TEOREMA 5.13 Si $Y_1, Y_2, \ldots, Y_k$ tienen una distribución binomial con parámetros y y $p_1, p_2, \ldots, p_k$, entonces

1. $E(Y_i) = np_i$, $V(Y_i) = np_iq_i$.
2. $\text{Cov}(Y_s, Y_t) = -np_sp_t$, si $s \neq t$.

Demostración La distribución marginal de Y_i se puede usar para obtener la media y varianza. Recuerde que Y_i puede ser interpretada como el número de intentos que caen en la celda i. Imagine todas las celdas, excluyendo la i, combinadas en una sola celda grande. Entonces cada intento resultará en la celda i o en una celda que no sea la i, con probabilidades p_i y $1 - p_i$, respectivamente. Entonces, Y_i posee una distribución de probabilidad marginal binomial. En consecuencia,

$$E(Y_i) = np_i \quad \text{y} \quad V(Y_i) = np_iq_i, \quad \text{donde } q_i = 1 - p_i.$$

Los mismos resultados se pueden obtener al establecer los valores esperados y evaluar. Por ejemplo,

$$E(Y_1) = \sum_{y_1}\sum_{y_2}\cdots\sum_{y_k} y_1 \frac{n!}{y_1!y_2!\ldots y_k!} p_1^{y_1} p_2^{y_2}\ldots p_k^{y_k}.$$

Como ya hemos deducido el valor esperado y la varianza de Y_i, dejamos la sumatoria de este valor esperado para el lector interesado.

La demostración de la parte 2 usa el Teorema 5.12. Considere el experimento multinomial como una sucesión de n intentos independientes y defina, para $s \neq t$,

$$U_i = \begin{cases} 1, & \text{si el intento } i \text{ resulta en clase } s, \\ 0, & \text{de otro modo,} \end{cases}$$

y

$$W_i = \begin{cases} 1, & \text{si el intento } i \text{ resulta en clase } t, \\ 0, & \text{de otro modo.} \end{cases}$$

Entonces

$$Y_s = \sum_{i=1}^{n} U_i \quad \text{y} \quad Y_t = \sum_{j=1}^{n} W_j.$$

(Como $U_i = 1$ o 0 dependiendo de si el i-ésimo intento resultó en clase s, Y_s es simplemente la suma de una serie de números 0 y 1. Ocurre un 1 en la suma cada vez que observemos un artículo de la clase s y un 0 cada vez que observemos cualquier otra clase. Entonces, Y_s es simplemente el número de veces que se observe la clase s. Una interpretación similar se aplica a Y_t.)

Observe que U_i y W_i no pueden ser iguales a 1 (el i-ésimo artículo no puede estar simultáneamente en las clases s y t). Entonces, el producto U_iW_i siempre es igual a cero y $E(U_iW_i) = 0$. Los siguientes resultados nos permiten evaluar $\text{Cov}(Y_s, Y_t)$:

$$\begin{aligned}
E(U_i) &= p_s \\
E(W_j) &= p_t \\
\text{Cov}(U_i, W_j) &= 0, \qquad \text{si } i \neq j \text{ porque los intentos son independientes} \\
\text{Cov}(U_i, W_i) &= E(U_i W_i) - E(U_i)E(W_i) = 0 - p_s p_t
\end{aligned}$$

Del Teorema 5.12 tenemos entonces

$$\begin{aligned}
\text{Cov}(Y_s, Y_t) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(U_i, W_j) \\
&= \sum_{i=1}^{n} \text{Cov}(U_i, W_i) + \sum_{i \neq j}\sum \text{Cov}(U_i, W_j) \\
&= \sum_{i=1}^{n} (-p_s p_t) + \sum_{i \neq j}\sum 0 = -np_s p_t.
\end{aligned}$$

La covarianza es negativa, lo que ya se esperaba, porque un gran número de resultados en la celda s forzaría al número de la celda t a ser pequeño.

Los problemas inferenciales asociados con el experimento multinomial se analizarán más adelante.

Ejercicios

5.119 Un experimento de aprendizaje requiere que una rata corra por un laberinto (una red de pasillos) hasta que localice una de tres posibles salidas. La salida 1 presenta una recompensa de alimento, no así las salidas 2 y 3. (Si la rata finalmente selecciona la salida 1 casi siempre, puede tener lugar el aprendizaje.) Denote con Y_i el número de veces que la salida i es seleccionada en corridas sucesivas. Para lo siguiente, suponga que la rata escoge una salida aleatoriamente en cada corrida.

- **a** Encuentre la probabilidad de que $n = 6$ corridas resulte en $Y_1 = 3$, $Y_2 = 1$ y $Y_3 = 2$.
- **b** Para n general, encuentre $E(Y_1)$ y $V(Y_1)$.
- **c** Encuentre $\text{Cov}(Y_2, Y_3)$ para n general.
- **d** Para comprobar la preferencia de la rata entre las salidas 2 y 3, podemos buscar en $Y_2 - Y_3$. Encuentre $E(Y_2 - Y_3)$ y $V(Y_2 - Y_3)$ para n general.

5.120 Una muestra de tamaño n se selecciona de un gran lote de artículos en los que una proporción p_1 contiene exactamente un defecto y una proporción p_2 contiene más de un defecto (con $p_1 + p_2 < 1$). El costo de reparar los artículos defectuosos de la muestra es $C = Y_1 + 3Y_2$, donde Y_1 denota el número de artículos con un defecto y Y_2 denota el número con dos o más defectos. Encuentre el valor esperado y la varianza de C.

5.121 Consulte el Ejercicio 5.117. Suponga que el número N de lagartos de la población es muy grande, con $p_1 = .3$ y $p_2 = .1$.

a Encuentre la probabilidad de que, en una muestra de cinco lagartos, $Y_1 = 2$ y $Y_2 = 1$.

b Si $n = 5$, encuentre $E\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right)$ y $V\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right)$.

5.122 Los pesos de una población de ratones alimentada con cierta dieta desde su nacimiento se supone que están normalmente distribuidos con $\mu = 100$ y $\sigma = 20$ (mediciones en gramos). Suponga que una muestra aleatoria de $n = 4$ ratones se toma de esta población. Encuentre la probabilidad de que

a exactamente dos ratones pesen entre 80 y 100 gramos y exactamente uno de ellos pese más de 100 gramos,

b los cuatro ratones pesen más de 100 gramos.

5.123 El National Fire Incident Reporting Service informó que, de los incendios residenciales, 73% son en casas familiares, 20% en departamentos y 7% en otros tipos de viviendas. Si se informa de cuatro incendios residenciales en un solo día, ¿cuál es la probabilidad de que dos sean en casas familiares, uno sea en un departamento y uno en otro tipo de vivienda?

5.124 El costo típico de daños causados por un incendio en una casa familiar es de \$20,000. Los costos comparables de un incendio en un departamento y en otros tipos de vivienda son \$10,000 y \$2000, respectivamente. Si cuatro incendios se reportan de manera independiente, use la información del Ejercicio 5.123 para hallar

a el costo total esperado de daños,

b la varianza del costo total de daños.

5.125 Cuando se inspeccionan aviones comerciales, las grietas en alas se reportan como no existentes, detectables o críticas. La historia de una flota particular indica que 70% de los aviones inspeccionados no tienen grietas en las alas, 25% tienen grietas detectables en las alas y 5% tienen grietas críticas en las alas. Se seleccionan aleatoriamente cinco aviones. Encuentre la probabilidad de que

a uno tenga una grieta crítica, dos tengan grietas detectables y dos no tengan grietas,

b al menos un avión tenga grietas críticas.

5.126 Un lote grande de artículos manufacturados contiene 10% con exactamente un defecto, 5% con más de un defecto y el resto sin defectos. Diez artículos se seleccionan aleatoriamente de entre este lote para su venta. Si Y_1 denota el número de artículos con un defecto y Y_2 el número con más de un defecto, los costos de reparación son $Y_1 + 3Y_2$. Encuentre la media y la varianza de los costos de reparación.

5.127 Consulte el Ejercicio 5.126. Denote con Y el número de artículos de entre los diez que contienen al menos un defecto. Encuentre la probabilidad de que Y

a sea igual a 2,

b sea al menos 1.

5.10 Distribución normal bivariante (opcional)

Ningún análisis de distribuciones de probabilidad multivariante estaría completo sin una referencia a la distribución normal multivariante, que es la piedra angular de mucha de la teoría moderna de estadística. En general, la función de densidad normal multivariante se define para

k variables aleatorias continuas, $Y_1, Y_2, \ldots, Y_k$. Debido a su complejidad, presentaremos sólo la función de densidad bivariante ($k = 2$):

$$f(y_1, y_2) = \frac{e^{-Q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \qquad -\infty < y_1 < \infty, \ -\infty < y_2 < \infty,$$

donde

$$Q = \frac{1}{1-\rho^2}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right].$$

La función de densidad normal bivariante es una función de cinco parámetros: μ_1, μ_2, σ_1^2, σ_2^2 y ρ. La elección de la notación empleada para estos parámetros no es casual. En el Ejercicio 5.128, usted demostrará que las distribuciones marginales de Y_1 y Y_2 son normales con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Con un poco de integración un tanto tediosa, podemos demostrar que $\text{Cov}(Y_1, Y_2) = \rho\sigma_1\sigma_2$.

Si $\text{Cov}(Y_1, Y_2) = 0$, o bien, lo que es equivalente, si $\rho = 0$, entonces

$$f(y_1, y_2) = g(y_1)h(y_2),$$

donde $g(y_1)$ es una función no negativa sólo de y_1 y $h(y_2)$ es una función no negativa sólo de y_2. Por tanto, si $\rho = 0$, el Teorema 5.5 implica que Y_1 y Y_2 son independientes. Recuerde que cero covarianza para dos variables aleatorias por lo general no implica independencia. No obstante, si Y_1 y Y_2 tienen una distribución normal bivariante, son independientes si y sólo si su covarianza es cero.

La expresión para la función de densidad conjunta, $k > 2$, se expresa con más facilidad usando álgebra de matrices. Un análisis del caso general se puede hallar en la bibliografía citada al final de este capítulo.

Ejercicios

***5.128** Considere que Y_1 y Y_2 tienen una distribución normal bivariante.

a Demuestre que la distribución marginal de Y_1 es normal con media μ_1 y varianza σ_1^2.

b ¿Cuál es la distribución marginal de Y_2?

***5.129** Considere que Y_1 y Y_2 tienen una distribución normal bivariante. Demuestre que la distribución condicional de Y_1 dado que $Y_2 = y_2$ es normal con media $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)$ y varianza $\sigma_1^2(1 - \rho^2)$.

***5.130** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$ para $i = 1, 2, \ldots, n$. Sean

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{y} \quad U_2 = \sum_{i=1}^{n} b_i Y_i,$$

donde $a_1, a_2, \ldots, a_n$ y $b_1, b_2, \ldots, b_n$ son constantes. Se dice que U_1 y U_2 son ortogonales si $\text{Cov}(U_1, U_2) = 0$.

a Demuestre que U_1 y U_2 son ortogonales si y sólo si $\sum_{i=1}^{n} a_i b_i = 0$.

b Suponga, además, que $Y_1, Y_2, \ldots, Y_n$ tienen una distribución normal multivariante. Entonces U_1 y U_2 tienen una distribución normal bivariante. Demuestre que U_1 y U_2 son independientes si son ortogonales.

***5.131** Sean Y_1 y Y_2 variables aleatorias independientes normalmente distribuidas con medias μ_1 y μ_2, respectivamente y varianzas $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

a Demuestre que Y_1 y Y_2 tienen distribución normal bivariante con $\rho = 0$.

b Considere $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$. Use el resultado del Ejercicio 5.130 para demostrar que U_1 y U_2 tienen una distribución normal bivariante y que U_1 y U_2 son independientes.

***5.132** Consulte el Ejercicio 5.131. ¿Cuáles son las distribuciones marginales de U_1 y U_2?

5.11 Valores esperados condicionales

La Sección 5.3 contiene un examen de funciones de probabilidad condicional y funciones de densidad condicional que ahora relacionaremos con valores esperados condicionales. Éstos se definen en la misma forma que los valores esperados univariantes, excepto que las densidades condicionales y las funciones de probabilidad se usan en lugar de sus similares marginales.

DEFINICIÓN **5.13**

Si Y_1 y Y_2 son dos variables aleatorias cualesquiera, el *valor esperado condicional* de $g(Y_1)$, dado que $Y_2 = y_2$, se define que es

$$E(g(Y_1) \mid Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1) f(y_1 \mid y_2)\, dy_1$$

si Y_1 y Y_2 son continuas conjuntamente y

$$E(g(Y_1) \mid Y_2 = y_2) = \sum_{\text{toda } y_1} g(y_1) p(y_1 \mid y_2)$$

si Y_1 y Y_2 son discretas conjuntamente.

EJEMPLO **5.31** Consulte las variables aleatorias Y_1 y Y_2 del Ejemplo 5.8, donde la función de densidad conjunta está dada por

$$f(y_1, y_2) = \begin{cases} 1/2, & 0 \le y_1 \le y_2 \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre el valor esperado condicional de la cantidad de ventas, Y_1, dado que $Y_2 = 1.5$.

Solución En el Ejemplo 5.8 determinamos que si $0 < y_2 \le 2$,

$$f(y_1 \mid y_2) = \begin{cases} 1/y_2, & 0 < y_1 \le y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Así, de acuerdo con la definición 5.13, para cualquier valor de y_2 tal que $0 < y_2 \le 2$,

$$\begin{aligned} E(Y_1 \mid Y_2 = y_2) &= \int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1 \\ &= \int_0^{y_2} y_1 \left(\frac{1}{y_2}\right) dy_1 = \frac{1}{y_2}\left(\left.\frac{y_1^2}{2}\right]_0^{y_2}\right) = \frac{y_2}{2}. \end{aligned}$$

En vista de que estamos interesados en el valor $y_2 = 1.5$, se deduce que $E(Y_1|Y_2 = 1.5) = 1.5/2 = 0.75$. Esto es, si la máquina automática expendedora de bebidas contiene 1.5 galones al principio del día, la cantidad esperada por venderse ese día es 0.75 galones. ■

En general, el valor esperado condicional de Y_1 dada $Y_2 = y_2$ es una función de y_2. Si ahora hacemos variar Y_2 en todos sus posibles valores, podemos considerar el valor esperado condicional $E(Y_1|Y_2)$ como una función de la variable aleatoria Y_2. En el Ejemplo 5.31 obtuvimos $E(Y_1|Y_2 = y_2) = y_2/2$. Se deduce que $E(Y_1|Y_2) = Y_2/2$. Como $E(Y_1|Y_2)$ es una función de la variable aleatoria Y_2, también es una variable aleatoria; y como tal, tiene media y varianza. Consideramos la media de esta variable aleatoria en el Teorema 5.14 y la varianza en el Teorema 5.15.

TEOREMA 5.14 Si Y_1 y Y_2 son dos variables aleatorias, entonces

$$E(Y_1) = E[E(Y_1 \mid Y_2)],$$

donde en el lado derecho de la ecuación el valor esperado interior es con respecto a la distribución condicional de Y_1 dada Y_2 y el valor esperado exterior es con respecto a la distribución de Y_2.

Demostración Suponga que Y_1 y Y_2 son continuas conjuntamente con función de densidad conjunta $f(y_1, y_2)$ y densidades marginales $f_1(y_1)$ y $f_2(y_2)$, respectivamente. Entonces

$$\begin{aligned}
E(Y_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2) f_2(y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} \left\{\int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1\right\} f_2(y_2)\, dy_2 \\
&= \int_{-\infty}^{\infty} E(Y_1 \mid Y_2 = y_2) f_2(y_2)\, dy_2 = E\left[E(Y_1 \mid Y_2)\right].
\end{aligned}$$

La demostración es semejante para el caso discreto.

EJEMPLO 5.32 Un programa de control de calidad para una línea de ensamble comprende el muestreo de $n = 10$ artículos terminados por día y contar el número de artículos defectuosos, Y. Si p denota la probabilidad de observar uno defectuoso, entonces Y tiene una distribución binomial, suponiendo que un número grande de artículos son producidos por la línea. Pero p varía diariamente y se supone que tiene una distribución uniforme en el intervalo de 0 a $1/4$. Encuentre el valor esperado de Y.

Solución Del Teorema 5.14 sabemos que $E(Y) = E[E(Y|p)]$. Para una p dada, Y tiene una distribución binomial, de ahí que $E(Y|p) = np$. Por tanto,

$$E(Y) = E[E(Y|p)] = E(np) = nE(p) = n\left(\frac{1/4 - 0}{2}\right) = \frac{n}{8},$$

y para $n = 10$

$$E(Y) = 10/8 = 1.25.$$

A la larga, esta política de inspección va a promediar 1.25 artículos defectuosos por día. ■

La varianza condicional de Y_1 dada $Y_2 = y_2$ está definida por analogía con una varianza ordinaria, de nuevo utilizando la densidad condicional o función de probabilidad de Y_1 dada $Y_2 = y_2$ en lugar de la densidad ordinaria o función de probabilidad de Y_1. Esto es,

$$V(Y_1 \mid Y_2 = y_2) = E(Y_1^2 \mid Y_2 = y_2) - [E(Y_1 \mid Y_2 = y_2)]^2 .$$

Como en el caso de la media condicional, la varianza condicional es una función de y_2. Si dejamos que Y_2 tome todos sus valores posibles, podemos definir $V(Y_1|Y_2)$ como una variable aleatoria que es una función de Y_2. Específicamente, si $g(y_2) = V(Y_1|Y_2 = y_2)$ es una función particular del valor observado y_2, entonces $g(Y_2)= V(Y_1|Y_2)$ es *la misma función* de la variable aleatoria, Y_2. El valor esperado de $V(Y_1|Y_2)$ es útil para calcular la varianza de Y_1, como se detalla en el Teorema 5.15.

TEOREMA 5.15 Si Y_1 y Y_2 representan variables aleatorias, entonces

$$V(Y_1) = E\left[V(Y_1 \mid Y_2)\right] + V\left[E(Y_1 \mid Y_2)\right].$$

Demostración Como se indicó antes, $V(Y_1|Y_2)$ está dada por

$$V(Y_1 \mid Y_2) = E(Y_1^2 \mid Y_2) - \left[E(Y_1 \mid Y_2)\right]^2$$

y

$$E\left[V(Y_1 \mid Y_2)\right] = E\left[E(Y_1^2 \mid Y_2)\right] - E\left\{\left[E(Y_1 \mid Y_2)\right]^2\right\}.$$

Por definición,

$$V\left[E(Y_1 \mid Y_2)\right] = E\left\{\left[E(Y_1 \mid Y_2)\right]^2\right\} - \left\{E\left[E(Y_1 \mid Y_2)\right]\right\}^2 .$$

La varianza de Y_1 es

$$\begin{aligned}
V(Y_1) &= E\left[Y_1^2\right] - \left[E(Y_1)\right]^2 \\
&= E\left\{E\left[Y_1^2 \mid Y_2\right]\right\} - \left\{E\left[E(Y_1 \mid Y_2)\right]\right\}^2 \\
&= E\left\{E\left[Y_1^2 \mid Y_2\right]\right\} - E\left\{\left[E(Y_1 \mid Y_2)\right]^2\right\} + E\left\{\left[E(Y_1 \mid Y_2)\right]^2\right\} \\
&\quad - \left\{E\left[E(Y_1 \mid Y_2)\right]\right\}^2 \\
&= E\left[V(Y_1 \mid Y_2)\right] + V\left[E(Y_1 \mid Y_2)\right].
\end{aligned}$$

EJEMPLO **5.33** Consulte el Ejemplo 5.32. Encuentre la varianza de Y.

Solución Del Teorema 5.15 sabemos que

$$V(Y_1) = E\big[V(Y_1 \mid Y_2)\big] + V\big[E(Y_1 \mid Y_2)\big].$$

Para una p dada, Y tiene una distribución binomial, y en consecuencia $E(Y|p) = np$ y $V(Y|p) = npq$. Por tanto,

$$\begin{aligned} V(Y) &= E\big[V(Y \mid p)\big] + V\big[E(Y \mid p)\big] \\ &= E(npq) + V(np) = nE\,[p(1-p)] + n^2V(p). \end{aligned}$$

Como p está uniformemente distribuida en el intervalo (0, 1/4) y $E(p^2) = V(p) + [E(p)]^2$, se deduce que

$$E(p) = \frac{1}{8}, \qquad V(p) = \frac{(1/4-0)^2}{12} = \frac{1}{192}, \qquad E(p^2) = \frac{1}{192} + \frac{1}{64} = \frac{1}{48}.$$

Así que,

$$\begin{aligned} V(Y) &= nE\,[p(1-p)] + n^2V(p) = n\big[E(p) - E(p^2)\big] + n^2V(p) \\ &= n\left(\frac{1}{8} - \frac{1}{48}\right) + n^2\left(\frac{1}{192}\right) = \frac{5n}{48} + \frac{n^2}{192}, \end{aligned}$$

y para $n = 10$,

$$V(Y) = 50/48 + 100/192 = 1.5625.$$

Por tanto, la desviación estándar de Y es $\sigma = \sqrt{1.5625} = 1.25$. ■

La media y varianza de Y calculada en los Ejemplos 5.32 y 5.33 podrían comprobarse determinando la función de probabilidad incondicional de Y y calcular $E(Y)$ y $V(Y)$ directamente. Para hacerlo, necesitaríamos hallar la distribución conjunta de Y y p. De esta distribución conjunta se puede obtener la función de probabilidad marginal de Y y determinar $E(Y)$ al evaluar $\sum_y yp(y)$. La varianza se puede determinar en la forma usual, de nuevo usando la función de probabilidad marginal de Y. En los Ejemplos 5.32 y 5.33, evitamos trabajar directamente con las distribuciones conjunta y marginal. Los Teoremas 5.14 y 5.15 nos permitieron un cálculo mucho más rápido de la media y la varianza deseadas. Como siempre, la media y la varianza de una variable aleatoria pueden combinarse con el teorema de Tchebysheff para obtener límites para probabilidades, cuando la distribución de la variable sea desconocida o difícil de deducir.

En los Ejemplos 5.32 y 5.33 encontramos una situación en la que la distribución de una variable aleatoria (Y = el número de artículos defectuosos) se dio *condicionalmente* para posibles valores de una cantidad p que podría variar de un día a otro. El hecho de que p variara se tomó en cuenta al asignar una distribución de probabilidad a esta variable. Éste es un ejemplo de un modelo *jerárquico*. En estos modelos la distribución de una variable de interés, por ejemplo, Y, está *condicionada* al valor de un "parámetro" θ. La incertidumbre alrededor del valor real de θ se modela al asignarle una distribución de probabilidad. Una vez que especificamos la distribución condicional de Y dada θ *y* la distribución marginal de θ, la distribución conjunta de Y y θ se obtiene al multiplicar la condicional por la marginal. La distribución marginal de Y se obtiene entonces de la distribución conjunta mediante una integración o suma que incluya los posibles valores de θ. Los resultados de esta sección se pueden usar para encontrar $E(Y)$ y $V(Y)$ *sin que sea necesario* determinar esta distribución marginal. Otros ejemplos de modelos jerárquicos están contenidos en los Ejercicios 5.136, 5.138, 5.141 y 5.142.

Ejercicios

5.133 En el Ejercicio 5.9 determinamos que

$$f(y_1, y_2) = \begin{cases} 6(1 - y_2), & 0 \leq y_1 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

es una función de densidad de probabilidad conjunta válida.

a Encuentre $E(Y_1|Y_2 = y_2)$.

b Use la respuesta que obtuvo en el inciso a para hallar $E(Y_1)$. (Compare esto con la respuesta hallada en el Ejercicio 5.77.)

5.134 En los Ejemplos 5.32 y 5.33 determinamos que si Y es el número de artículos defectuosos, $E(Y) = 1.25$ y $V(Y) = 1.5625$. ¿Es probable que, en cualquier día determinado, Y sea mayor que 6?

5.135 En el Ejercicio 5.41 consideramos un programa de control de calidad que exige seleccionar en forma aleatoria tres artículos de entre la producción diaria (que se supone es grande) de cierta máquina y observar el número de artículos defectuosos. La proporción p de artículos defectuosos producidos por la máquina varía diariamente y tiene una distribución uniforme en el intervalo (0, 1). Encuentre

a el número esperado de artículos defectuosos observado entre los tres artículos muestreados.

b la varianza del número de artículos defectuosos de entre los tres muestreados.

5.136 En el Ejercicio 5.42 se supo que el número de defectos por yarda en cierta tela, Y, tenía una distribución de Poisson con parámetro λ. Se supuso que el parámetro λ era una variable aleatoria con función de densidad dada por

$$f(\lambda) = \begin{cases} e^{-\lambda}, & \lambda \geq 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el número esperado de defectos por yarda hallando primero el valor condicional de Y para una λ dada.

b Encuentre la varianza de Y.

c ¿Es probable que Y exceda de 9?

5.137 En el Ejercicio 5.38 supusimos que Y_1, el peso de un artículo a granel abastecido por un proveedor, tenía una distribución uniforme en el intervalo (0, 1). La variable aleatoria Y_2 denotaba el peso del artículo vendido y se suponía que tenía una distribución uniforme en el intervalo $(0, y_1)$, donde y_1 era un valor específico de Y_1. Si el proveedor abasteció 3/4 de tonelada, ¿qué cantidad podría esperar vender durante la semana?

5.138 Suponga que Y denota el número de bacterias por centímetro cúbico en un líquido particular y que Y tiene una distribución de Poisson con parámetro λ. También suponga que λ varía de un lugar a otro y tiene una distribución gamma con parámetros α y β, donde α es un entero positivo. Si aleatoriamente seleccionamos un lugar, ¿cuál es

a el número esperado de bacterias por centímetro cúbico?,

b la desviación estándar del número de bacterias por centímetro cúbico?

5.139 Suponga que una compañía ha determinado que el número de trabajos por semana, N, varía de una semana a otra y tiene una distribución de Poisson con media λ. El número de horas para completar cada trabajo, Y_i, es una distribución gamma con parámetros α y β. El tiempo total para completar todos los trabajos en una semana es $T = \sum_{i=1}^{N} Y_i$. Observe que T es la suma de un número aleatorio de variables aleatorias. ¿Cuál es

a $E(T|N = n)$?,

b $E(T)$, el tiempo total esperado para completar todos los trabajos?

5.140 ¿Por qué es $E[V(Y_1|Y_2)] \leq V(Y_1)$?

5.141 Sea Y_1 que tiene una distribución exponencial con media λ y la densidad condicional de Y_2 dado que $Y_1 = y_1$ es

$$f(y_2 \mid y_1) = \begin{cases} 1/y_1, & 0 \leq y_2 \leq y_1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre $E(Y_2)$ y $V(Y_2)$, la media y la varianza incondicional de Y_2.

5.142 Suponga que Y tiene una distribución binomial con parámetros n y p pero que p varía de un día a otro de acuerdo con una distribución beta con parámetros α y β. Demuestre que

a $E(Y) = n\alpha/(\alpha + \beta)$.

b $V(Y) = \dfrac{n\alpha\beta\,(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

***5.143** Si Y_1 y Y_2 son variables aleatorias independientes, cada una teniendo una distribución normal con media 0 y varianza 1, encuentre la función generadora de momento de $U = Y_1Y_2$. Use esta función generadora de momento para hallar $E(U)$ y $V(U)$. Compruebe el resultado al evaluar $E(U)$ y $V(U)$ directamente de las funciones de densidad para Y_1 y Y_2.

5.12 Resumen

El experimento multinomial (Sección 5.9) y su distribución de probabilidad multinomial asociada constituyen el tema de este capítulo. Casi todos los experimentos dan mediciones de muestra, $y_1, y_2, \ldots, y_k$, que pueden ser considerados como observaciones de k variables aleatorias. Las inferencias relacionadas con la estructura que generan las observaciones, es decir, las probabilidades de caer en las celdas $1, 2, \ldots, k$, están basadas en el conocimiento de las probabilidades asociadas con diversas muestras $(y_1, y_2, \ldots, y_k)$. Las distribuciones conjunta, marginal y condicional son conceptos esenciales para determinar las probabilidades de varios resultados muestrales.

Generalmente tomamos una muestra de n observaciones de una población, que son valores específicos de $Y_1, Y_2, \ldots, Y_n$. Muchas veces las variables aleatorias son independientes y tienen la misma distribución de probabilidad. En consecuencia, el concepto de independencia es útil para calcular la probabilidad de observar la muestra dada.

El objetivo de este capítulo fue expresar las ideas contenidas en los dos párrafos anteriores. La gran cantidad de detalles contenidos en el capítulo son esenciales para dar un sólido respaldo para un estudio de inferencia. Al mismo tiempo, el lector debe tener el cuidado de evitar dar demasiada importancia a los detalles; asegúrese de tener en mente la amplitud de los objetivos de la inferencia.

Bibliografía y lecturas adicionales

Hoel, P. G. 1984. *Introduction to Mathematical Statistics*, 5th ed. New York: Wiley.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill.

Myers, R. H. 2000. *Classical and Modern Regression with Applications*, 2d ed. Pacific Grove, CA:Duxbury Press.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Ejercicios complementarios

5.144 Demuestre el Teorema 5.9 cuando Y_1 y Y_2 son variables aleatorias discretas independientes.

5.145 Un técnico empieza un trabajo en el tiempo Y_1 que está uniformemente distribuido entre las 8:00 a.m. y las 8:15 a.m. El tiempo para completar el trabajo, Y_2, es una variable aleatoria independiente que está uniformemente distribuida entre 20 y 30 minutos. ¿Cuál es la probabilidad de que el trabajo se complete antes de las 8:30 a.m.?

5.146 El objetivo de una bomba está en el centro de un círculo con radio de 1 milla. La bomba cae en un punto seleccionado aleatoriamente dentro de ese círculo. Si la bomba destruye todo dentro de $1/2$ milla de su punto de caída, ¿cuál es la probabilidad de que el objetivo sea destruido?

5.147 Dos amigos se han de encontrar en la biblioteca. Cada uno de ellos, en forma independiente y aleatoria, selecciona una hora de llegada dentro del mismo periodo de una hora y conviene en esperar diez minutos al otro hasta que llegue. ¿Cuál es la probabilidad de que se encuentren?

5.148 Una comisión de tres personas va a ser seleccionada aleatoriamente de entre un grupo que contiene cuatro republicanos, tres demócratas y dos independientes. Denote con Y_1 y Y_2 los números de republicanos y demócratas, respectivamente, en la comisión.

a ¿Cuál es la distribución de probabilidad conjunta para Y_1 y Y_2?

b Encuentre las distribuciones marginales de Y_1 y Y_2.

c Encuentre $P(Y_1 = 1|Y_2 \geq 1)$.

5.149 Tengan Y_1 y Y_2 una función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \leq y_2 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre las funciones de densidad marginal de Y_1 y Y_2.

b Encuentre $P(Y_1 \leq 3/4|Y_2 \leq 1/2)$.

c Encuentre la función de densidad condicional de Y_1 dado que $Y_2 = y_2$.

d Encuentre $P(Y_1 \leq 3/4|Y_2 = 1/2)$.

5.150 Consulte el Ejercicio 5.149.

a Encuentre $E(Y_2|Y_1 = y_1)$.

b Use el Teorema 5.14 para hallar $E(Y_2)$.

c Encuentre $E(Y_2)$ directamente de la densidad marginal de Y_2.

5.151 La vida útil Y para un tipo de fusible tiene una distribución exponencial con una función de densidad dada por

$$f(y) = \begin{cases} (1/\beta)e^{-y/\beta}, & y \geq 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Si dos de estos fusibles tienen vidas independientes Y_1 y Y_2, encuentre su función de densidad de probabilidad conjunta.

b Uno de los fusibles del inciso a está en un sistema primario y el otro está en el sistema de respaldo que entra en servicio sólo si falla el sistema primario. Por tanto, la vida útil total efectiva de los dos fusibles es $Y_1 + Y_2$. Encuentre $P(Y_1 + Y_2 \leq a)$, donde $a > 0$.

5.152 En la producción de cierto tipo de cobre, dos tipos de polvo de cobre (A y B) se mezclan y se sinterizan (calientan) durante cierto tiempo. Para un volumen fijo de cobre sinterizado, el productor mide la proporción Y_1 del volumen debido al cobre sólido (algunos poros tendrán que llenarse de aire) y la proporción Y_2 de la masa sólida debida a cristales tipo A. Suponga que las densidades de probabilidad apropiadas para Y_1 y Y_2 son

$$f_1(y_1) = \begin{cases} 6y_1(1 - y_1), & 0 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

$$f_2(y_2) = \begin{cases} 3y_2^2, & 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La proporción del volumen muestral debida a cristales tipo A es entonces Y_1Y_2. Suponiendo que Y_1 y Y_2 sean independientes, encuentre $P(Y_1Y_2 \leq .5)$.

5.153 Suponga que el número de huevos puestos por cierto insecto tiene una distribución de Poisson con media λ. La probabilidad de que cualquier huevo incube es p. Suponga que los huevos se incuban de manera independiente unos de otros. Encuentre

a el valor esperado de Y, el número total de huevos que incuban,

b la varianza de Y.

5.154 En un estudio clínico de un nuevo medicamento formulado para reducir los efectos de la artritis reumatoide los investigadores encontraron que la proporción p de pacientes que responden de manera favorable al medicamento es una variable aleatoria que varía con cada lote del medicamento. Suponga que p tiene una función de densidad de probabilidad dada por

$$f(p) = \begin{cases} 12p^2(1 - p), & 0 \leq p \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Suponga que a n pacientes se les inyectan porciones del medicamento tomadas del mismo lote. Denote con Y el número que muestre una respuesta favorable.

a Encuentre la distribución de probabilidad incondicional de Y para n general.

b Calcule $E(Y)$ para $n = 2$.

5.155 Suponga que Y_1, Y_2 y Y_3 son variables aleatorias independientes con distribución χ^2, con ν_1, ν_2 y ν_3 grados de libertad respectivamente y que $W_1 = Y_1 + Y_2$ y $W_2 = Y_1 + Y_3$.

a En el Ejercicio 5.87, usted obtuvo la media y varianza de W_1. Encuentre Cov(W_1, W_2).

b Explique por qué esperaba que la respuesta al inciso a fuera positiva.

5.156 Consulte el Ejercicio 5.86. Suponga que Z es una variable aleatoria normal estándar y que Y es una variable aleatoria independiente con distribución χ^2 y ν grados de libertad.

a Defina $W = Z/\sqrt{Y}$. Encuentre Cov(Z, W). ¿Qué suposición necesita acerca del valor de ν?

b Con Z, Y y W como se indica líneas antes, encuentre Cov(Y, W).

c Una de las covarianzas de los incisos a y b es positiva, y la otra es cero. Explique por qué.

5.157 Un guardabosques que estudia pinos enfermos representa el número de árboles enfermos por acre, Y, como una variable aleatoria de Poisson con media λ. No obstante, λ cambia de una zona a otra y su comportamiento aleatorio es modelado por una distribución gamma. Esto es, para algún entero α,

$$f(\lambda) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}}\lambda^{\alpha-1}e^{-\lambda/\beta}, & \lambda > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la distribución de probabilidad incondicional para Y.

5.158 Una moneda tiene probabilidad p de caer de cara hacia arriba cuando se lanza al aire. En n lanzamientos independientes, $X_i = 1$ si el i-ésimo tiro resulta cara y $X_i = 0$ si el i-ésimo tiro resulta cruz. Entonces Y, el número de caras en los n lanzamientos, tiene una distribución binomial y puede representarse como $Y = \sum_{i=1}^{n} X_i$. Encuentre $E(Y)$ y $V(Y)$ usando el Teorema 5.12.

***5.159** La variable aleatoria binomial negativa Y se definió en la Sección 3.6 como el número del intento en el que ocurre el r-ésimo éxito, en una sucesión de intentos independientes con probabilidad constante p de éxito en cada intento. Denote con X_i una variable aleatoria definida como el número del intento en el que ocurre el i-ésimo éxito, para $i = 1, 2, \ldots, r$. Ahora defina

$$W_i = X_i - X_{i-1}, \qquad i = 1, 2, \ldots, r,$$

donde a X_0 se le asigna un valor de cero. Entonces puede escribir $Y = \sum_{i=1}^{r} W_i$. Observe que las variables aleatorias $W_1, W_2, \ldots, W_r$ tienen distribuciones geométricas idénticas y son mutuamente independientes. Aplique el Teorema 5.12 para demostrar que $E(Y) = r/p$ y $V(Y) = r(1-p)/p^2$.

5.160 Una caja contiene cuatro pelotas, numeradas del 1 al 4. Una de ellas se selecciona aleatoriamente de la caja. Sean

$$X_1 = 1 \text{ si se saca la pelota 1 o la pelota 2,}$$
$$X_2 = 1 \text{ si se saca la pelota 1 o la pelota 3,}$$
$$X_3 = 1 \text{ si se saca la pelota 1 o la pelota 4.}$$

En cualquier otro caso X_i es igual a cero. Demuestre que dos variables aleatorias cualesquiera X_1, X_2 y X_3 son independientes mientras que las tres juntas no lo son.

5.161 Suponga que debemos observar dos muestras aleatorias independientes: $Y_1, Y_2, \ldots, Y_n$ denota una muestra aleatoria de una distribución normal con media μ_1 y varianza σ_1^2; y $X_1, X_2, \ldots, X_n$ denota una muestra aleatoria de otra distribución normal con media μ_2 y varianza σ_2^2. Una aproximación para $\mu_1 - \mu_2$ está dada por $\overline{Y} - \overline{X}$, la diferencia entre las medias muestrales. Encuentre $E(\overline{Y} - \overline{X})$ y $V(\overline{Y} - \overline{X})$.

5.162 En el Ejercicio 5.65 usted determinó que para $-1 \le \alpha \le 1$, la función de densidad de probabilidad de (Y_1, Y_2) está dada por

$$f(y_1, y_2) = \begin{cases} [1 - \alpha\{(1 - 2e^{-y_1})(1 - 2e^{-y_2})\}]e^{-y_1 - y_2}, & 0 \le y_1, 0 \le y_2, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y es tal que las distribuciones marginales de Y_1 y Y_2 son exponenciales con media de 1. También demostró que Y_1 y Y_2 son independientes si y sólo si $\alpha = 0$. Proporcione dos densidades conjuntas específicas y diferentes que produzcan densidades marginales para Y_1 y Y_2 que sean exponenciales con media de 1.

***5.163** Consulte el Ejercicio 5.66. Si $F_1(y_1)$ y $F_2(y_2)$ son dos funciones de distribución, entonces para cualquier α, $-1 \le \alpha \le 1$,

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)[1 - \alpha\{1 - F_1(y_1)\}\{1 - F_2(y_2)\}]$$

es una función de distribución conjunta tal que Y_1 y Y_2 tienen funciones de distribución marginal $F_1(y_1)$ y $F_2(y_2)$, respectivamente.

a Si $F_1(y_1)$ y $F_2(y_2)$ son funciones de distribución asociadas con variables aleatorias exponencialmente distribuidas con media de 1, demuestre que la *función de densidad conjunta* de Y_1 y Y_2 es la proporcionada en el Ejercicio 5.162.

b Si $F_1(y_1)$ y $F_2(y_2)$ son funciones de distribución asociadas con variables aleatorias uniformes (0, 1), para cualquier α, $-1 \le \alpha \le 1$, evalúe $F(y_1, y_2)$.

c Encuentre las *funciones de densidad conjunta* asociadas con las funciones de distribución que encontró en el inciso b.

d Obtenga dos densidades conjuntas específicas y diferentes tales que las distribuciones marginales de Y_1 y Y_2 sean uniformes en el intervalo (0, 1).

***5.164** Sean X_1, X_2 y X_3 variables aleatorias, ya sea continuas o discretas. La función generadora de momentos conjunta de X_1, X_2 y X_3 está definida por

$$m(t_1, t_2, t_3) = E(e^{t_1 X_1 + t_2 X_2 + t_3 X_3}).$$

a Demuestre que $m(t, t, t)$ representa la función generadora de momentos de $X_1 + X_2 + X_3$.

b Demuestre que $m(t, t, 0)$ representa la función generadora de momentos de $X_1 + X_2$.

c Demuestre que

$$\left.\frac{\partial^{k_1+k_2+k_3} m(t_1, t_2, t_3)}{\partial t_1^{k_1} \partial t_2^{k_2} \partial t_3^{k_3}}\right]_{t_1=t_2=t_3=0} = E\left(X_1^{k_1} X_2^{k_2} X_3^{k_3}\right).$$

***5.165** Sean X_1, X_2 y X_3 que tienen distribución multinomial con función de probabilidad

$$p(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}, \qquad \sum_{i=1}^{n} x_i = n.$$

Use los resultados del Ejercicio 5.164 para responder lo siguiente:

a Encuentre la función generadora de momentos conjunta de X_1, X_2 y X_3.

b Utilice la respuesta al inciso a para demostrar que la distribución marginal de X_1 es binomial con parámetro p_1.

c Utilice la función generadora de momentos conjunta para hallar $\text{Cov}(X_1, X_2)$.

***5.166** Una caja contiene N_1 bolas blancas, N_2 bolas negras y N_3 bolas rojas ($N_1 + N_2 + N_3 = N$). Se toma de la caja una muestra aleatoria de n bolas (sin restitución). Si Y_1, Y_2 y Y_3 representan el número de bolas blancas, negras y rojas observadas en la muestra, respectivamente, determine el coeficiente de correlación para Y_1 y Y_2. (Sea $p_i = N_i/N$, para $i = 1, 2, 3$.)

***5.167** Sean Y_1 y Y_2 variables aleatorias distribuidas conjuntamente con varianzas finitas.

a Demuestre que $[E(Y_1Y_2)]^2 \leq E(Y_1^2)E(Y_2^2)$. [*Sugerencia:* observe que $E[(tY_1 - Y_2)^2] \geq 0$ para cualquier número real t, o bien, de manera equivalente,

$$t^2E(Y_1^2) - 2tE(Y_1Y_2) + E(Y_2^2) \geq 0.$$

Ésta es una expresión cuadrática de la forma $At^2 + Bt + C$ y como es no negativa, se tiene que $B^2 - 4AC \leq 0$. La desigualdad anterior se deduce directamente.]

b Denote con ρ el coeficiente de correlación de Y_1 y Y_2. Usando la desigualdad del inciso a, demuestre que $\rho^2 \leq 1$.

CAPÍTULO 6

Funciones de variables aleatorias

6.1 Introducción

Como ya indicamos en el Capítulo 1, el objetivo de la estadística es hacer inferencias acerca de una población con base en información contenida en una muestra tomada de esa población. Cualquier inferencia verdaderamente útil debe ser acompañada por una medida de bondad asociada. Cada uno de los temas que se estudiaron en los capítulos anteriores desempeña un papel en el desarrollo de la inferencia estadística. No obstante, ninguno de los temas examinados hasta aquí está relacionado con el objetivo de la estadística en forma tan cercana como el estudio de las distribuciones de funciones de variables aleatorias. Esto se debe a que todas las cantidades empleadas para calcular parámetros poblacionales o tomar decisiones acerca de una población, son funciones de las n observaciones aleatorias que aparecen en una muestra.

Para ilustrar, considere el problema de estimar una media poblacional, μ. De manera intuitiva tomamos una muestra aleatoria de n observaciones, $y_1, y_2, \ldots, y_n$, de la población y utilizamos la media muestral

$$\overline{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

como estimación de μ. ¿Qué tan buena es esta estimación? La respuesta depende del comportamiento de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ y su efecto en la distribución de $\overline{Y} = (1/n)\sum_{i=1}^{n} Y_i$.

El *error de estimación*, la diferencia entre la estimación y el parámetro estimado (para nuestro ejemplo, la diferencia entre $\overline{y}$ y μ) constituyen una medida de la bondad de una estimación. Como $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias, al repetir el muestreo, $\overline{Y}$ también es una variable aleatoria (y una función de las n variables $Y_1, Y_2, \ldots, Y_n$). Por tanto, no podemos estar seguros de que el error de estimación sea menor que un valor específico, B, por ejemplo. Sin embargo, si pudiéramos determinar la distribución de probabilidad del estimador $\overline{Y}$, la utilizaríamos para determinar la *probabilidad* de que el error de estimación sea menor o igual a B.

Para determinar la distribución de probabilidad para una función de n variables aleatorias, $Y_1, Y_2, \ldots, Y_n$, debemos calcular la distribución de probabilidad conjunta para las variables aleatorias mismas. Por lo general suponemos que las observaciones se obtienen mediante muestreo aleatorio, como se definió en la Sección 2.12. Vimos en la Sección 3.7 que el muestreo aleatorio a partir de una población finita (muestreo sin restitución) resulta en intentos dependientes, que se hacen esencialmente independientes si la población es grande cuando se compara con el tamaño de la muestra.

En todo el resto de este libro supondremos que las poblaciones son grandes en comparación con el tamaño muestral y, en consecuencia, que las variables aleatorias obtenidas a través de muestreo aleatorio son, de hecho, independientes entre sí. Entonces, en el caso discreto, la función de probabilidad conjunta para las variables $Y_1, Y_2, \ldots, Y_n$, seleccionadas de la misma población, está dada por

$$p(y_1, y_2, \ldots, y_n) = p(y_1)p(y_2)\cdots p(y_n).$$

En el caso continuo, la función de densidad conjunta es

$$f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2)\cdots f(y_n).$$

El enunciado "$Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una población con densidad $f(y)$" significa que las variables aleatorias son independientes con función de densidad común $f(y)$.

6.2 Determinación de la distribución de probabilidad de una función de variables aleatorias

En esta sección presentaremos tres métodos para determinar la distribución de probabilidad para una función de variables aleatorias y un cuarto método para determinar la distribución *conjunta* de varias funciones de variables aleatorias. Cualquiera de estos métodos se puede emplear para determinar la distribución de una función dada de las variables, pero por lo general, uno de los métodos lleva a un proceso más sencillo que los otros. El método que funciona "mejor" varía de una aplicación a otra. En consecuencia, el conocimiento de los primeros tres métodos es deseable. El cuarto método se presenta en la Sección 6.6 (opcional). Aun cuando los tres primeros métodos se estudiarán por separado en las siguientes tres secciones, aquí damos un breve resumen de cada uno de ellos.

Considere las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ y una función $U(Y_1, Y_2, \ldots, Y_n)$, denotada simplemente como U. Entonces tres de los métodos para determinar la distribución de probabilidad de U son los siguientes:

1. Método de las funciones de distribución: se emplea generalmente cuando las Y tienen distribuciones continuas. Primero se determina la función de distribución para U, $F_U(u) = P(U \leq u)$, usando los métodos que estudiamos en el Capítulo 5. Para hacerlo, debemos determinar la región en el espacio $y_1, y_2, \ldots, y_n$ para la cual $U \leq u$ y entonces se determina $P(U \leq u)$ al integrar $f(y_1, y_2, \ldots, y_n)$ para esta región. La función de densidad para U se obtiene entonces al diferenciar la función de distribución, $F_U(u)$. En la Sección 6.3 presentaremos una explicación detallada de este procedimiento.
2. Método de las transformaciones: si nos dan la función de densidad de una variable aleatoria Y, el método de transformaciones resulta en una expresión general para la densidad de $U = h(Y)$, donde $h(y)$ es una función creciente o decreciente. Entonces, si Y_1 y Y_2 tienen una distribución bivariante, podemos usar el resultado univariante explicado antes para hallar la densidad conjunta de Y_1 y $U = h(Y_1, Y_2)$. Al integrar para y_1, encontramos la función de densidad de probabilidad marginal de U, que es nuestro objetivo. Este método se ilustra en la Sección 6.4.
3. Método de las funciones generadoras de momento: está basado en el teorema de unicidad, el Teorema 6.1, el cual expresa que si dos variables aleatorias tienen funciones generadoras de momento idénticas, las dos poseen las mismas distribuciones de probabilidad. Para usar este método es necesario determinar la función generadora de momento de U y compararla con las funciones generadoras de momento para las variables aleatorias discretas y continuas comunes deducidas en los Capítulos 3 y 4. Si aquélla es idéntica a una de estas funciones generadoras de momento, la distribución de probabilidad de U se puede identificar debido a la unicidad del teorema. Las aplicaciones del método de funciones generadoras de momento se presentarán en la Sección 6.5. Las funciones generadoras de probabilidad se pueden emplear en forma semejante al método de funciones generadoras de momento. Si el lector está interesado en su uso, vea la bibliografía al final del capítulo.

6.3 Método de las funciones de distribución

Ilustraremos el método de funciones de distribución con un ejemplo sencillo univariante. Si Y tiene función de densidad de probabilidad $f(y)$ y si U es alguna función de Y, entonces podemos calcular $F_U(u) = P(U \leq u)$ directamente al integrar $f(y)$ en la región para la cual $U \leq u$. La función de densidad de probabilidad para U se encuentra al derivar $F_U(u)$. El siguiente ejemplo ilustra el método.

EJEMPLO **6.1** Un proceso para refinar azúcar rinde hasta 1 tonelada de azúcar puro al día, pero la cantidad real producida, Y, es una variable aleatoria debido a descomposturas de máquinas y otros problemas. Suponga que Y tiene función de densidad dada por

$$f(y) = \begin{cases} 2y, & 0 \leq y \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

A la compañía se le paga a razón de \$300 por tonelada de azúcar refinada, pero también tiene un costo fijo general de \$100 por día. Por tanto, la utilidad diaria, en cientos de dólares, es $U = 3Y - 1$. Encuentre la función de densidad de probabilidad para U.

Solución Para utilizar el método de función de distribución, debemos hallar

$$F_U(u) = P(U \le u) = P(3Y - 1 \le u) = P\left(Y \le \frac{u+1}{3}\right).$$

Si $u < -1$, entonces $(u + 1)/3 < 0$ y, por tanto, $F_U(u) = P(Y \le (u + 1)/3) = 0$. También, si $u > 2$, entonces $(u + 1)/3 > 1$ y $F_U(u) = P(Y \le (u + 1)/3) = 1$. No obstante, si $-1 \le u \le 2$, la probabilidad se puede escribir como una integral de $f(y)$, y

$$P\left(Y \le \frac{u+1}{3}\right) = \int_{-\infty}^{(u+1)/3} f(y)dy = \int_{0}^{(u+1)/3} 2y\, dy = \left(\frac{u+1}{3}\right)^2.$$

(Observe que, cuando Y varía de 0 a 1, U varía de -1 a 2.) Entonces, la función de distribución de la variable aleatoria U está dada por

$$F_U(u) = \begin{cases} 0, & u < -1, \\ \left(\dfrac{u+1}{3}\right)^2, & -1 \le u \le 2, \\ 1, & u > 2, \end{cases}$$

y la función de densidad para U es

$$f_U(u) = \frac{dF_U(u)}{du} = \begin{cases} (2/9)(u+1), & -1 \le u < 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$ ■

En la situación bivariante, sean Y_1 y Y_2 variables aleatorias con densidad conjunta $f(y_1, y_2)$ y sea $U = h(Y_1, Y_2)$ una función de Y_1 y Y_2. Entonces, para todo punto (y_1, y_2), corresponde un valor y sólo uno de U. Si podemos precisar la región de valores (y_1, y_2) tal que $U \le u$, entonces la integral de la función de densidad conjunta $f(y_1, y_2)$ para esta región es igual a $P(U \le u) = F_U(u)$. Como antes, la función de densidad para U se puede obtener por derivación.

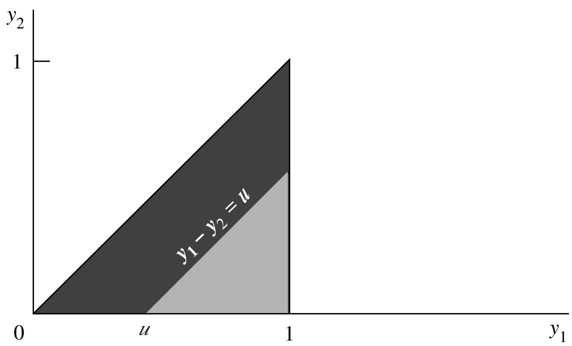
Ilustraremos estas ideas con dos ejemplos.

EJEMPLO **6.2** En el Ejemplo 5.4 consideramos las variables aleatorias Y_1 (la cantidad proporcional de gasolina abastecida al principio de semana) y Y_2 (la cantidad proporcional de gasolina vendida durante la semana). La función de densidad conjunta de Y_1 y Y_2 está dada por

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad de probabilidad para $U = Y_1 - Y_2$, la cantidad proporcional de gasolina remanente al final de la semana. Use la función de densidad de U para hallar $E(U)$.

FIGURA **6.1** Región en la cual $f(y_1, y_2)$ es positiva, Ejemplo 6.2



Solución La región en la cual $f(y_1, y_2)$ no es cero, así como la línea $y_1 - y_2 = u$, para un valor de u entre 0 y 1 se muestran en la Figura 6.1. Observe que cualquier punto (y_1, y_2) tal que $y_1 - y_2 \leq u$ se encuentra sobre la recta $y_1 - y_2 = u$.

Si $u < 0$, la recta $y_1 - y_2 = u$ tiene punto de intersección $-u < 0$ y $F_U(u) = P(Y_1 - Y_2 \leq u) = 0$. Cuando $u > 1$, la recta $y_1 - y_2 = u$ tiene punto de intersección $-u < -1$ y $F_U(u) = 1$. Para $0 \leq u \leq 1$, $F_U(u) = P(Y_1 - Y_2 \leq u)$ es la integral sobre la región sombreada oscura localizada arriba de la recta $y_1 - y_2 = u$. Como es más fácil integrar sobre la región triangular inferior, podemos escribir, para $0 \leq u \leq 1$,

$$\begin{aligned}
F_U(u) &= P(U \leq u) = 1 - P(U \geq u) \\
&= 1 - \int_u^1 \int_0^{y_1 - u} 3y_1 \, dy_2 \, dy_1 \\
&= 1 - \int_u^1 3y_1(y_1 - u) \, dy_1 \\
&= 1 - 3\left(\frac{y_1^3}{3} - \frac{uy_1^2}{2}\right)\Bigg]_u^1 \\
&= 1 - \left[1 - \frac{3}{2}(u) + \frac{u^3}{2}\right] \\
&= \frac{1}{2}(3u - u^3).
\end{aligned}$$

Resumiendo,

$$F_U(u) = \begin{cases} 0, & u < 0, \\ (3u - u^3)/2, & 0 \leq u \leq 1, \\ 1, & u > 1. \end{cases}$$

Una gráfica de $F_U(u)$ se muestra en la Figura 6.2(a).

Se deduce que

$$f_U(u) = \frac{dF_U(u)}{du} = \begin{cases} 3(1 - u^2)/2, & 0 \leq u \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La función de densidad $f_U(u)$ se grafica en la figura 6.2(b).

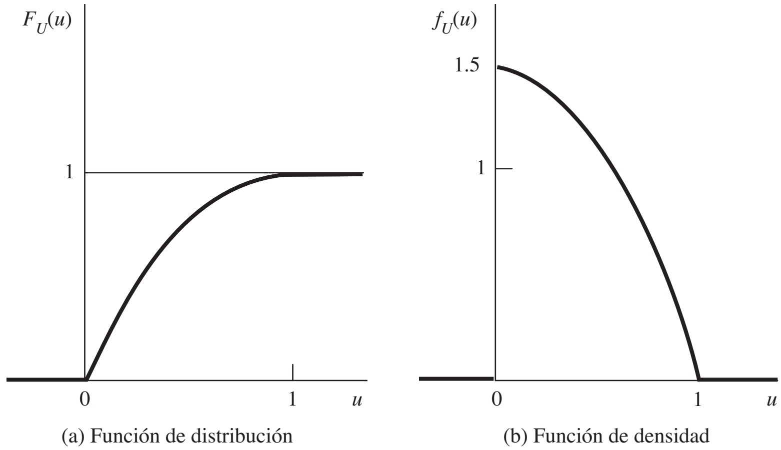


FIGURA **6.2** Funciones de distribución y densidad para el Ejemplo 6.2

Podemos usar esta función de densidad para hallar $E(U)$, porque

$$E(U) = \int_0^1 u\left(\frac{3}{2}\right)(1 - u^2)\,du = \frac{3}{2}\left(\frac{u^2}{2} - \frac{u^4}{4}\right)\Big]_0^1 = \frac{3}{8},$$

lo cual concuerda con el valor de $E(Y_1 - Y_2)$ encontrado en el Ejemplo 5.20 con el uso de métodos desarrollados en el Capítulo 5 para deducir el valor esperado de una función lineal de variables aleatorias. ■

EJEMPLO **6.3** Denote con (Y_1, Y_2) una muestra aleatoria de tamaño $n = 2$ proveniente de la distribución uniforme del intervalo (0, 1). Encuentre la función de densidad de probabilidad para $U = Y_1 + Y_2$.

Solución La función de densidad para cada Y_i es

$$f(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por tanto, como tenemos una muestra aleatoria, Y_1 y Y_2 son independientes, y

$$f(y_1, y_2) = f(y_1)f(y_2) = \begin{cases} 1, & 0 \le y_1 \le 1,\ 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Las variables aleatorias Y_1 y Y_2 tienen densidad diferente de cero en el cuadrado unitario, como se ve en la Figura 6.3. Deseamos determinar $F_U(u) = P(U \le u)$. El primer paso es localizar los puntos (y_1, y_2) que implican que $y_1 + y_2 \le u$. La forma más fácil de encontrar esta región es localizar los puntos que dividen las regiones $U \le u$ y $U > u$. Estos puntos se encuentran en la recta $y_1 + y_2 = u$.

Al graficar esta relación en la figura 6.3 y en forma arbitraria seleccionar y_2 como la variable dependiente, encontramos que la recta posee una pendiente igual a −1 y un punto de intersección y_2 igual a u. Los puntos asociados con $U < u$ están ya sea arriba o debajo de la recta y pueden ser determinados al probar puntos en cualquiera de los lados de la recta. Suponga que $u = 1.5$. Sea $y_1 = y_2 = 1/4$; entonces $y_1 + y_2 = 1/4 + 1/4 = 1/2$ y (y_1, y_2) satisface la

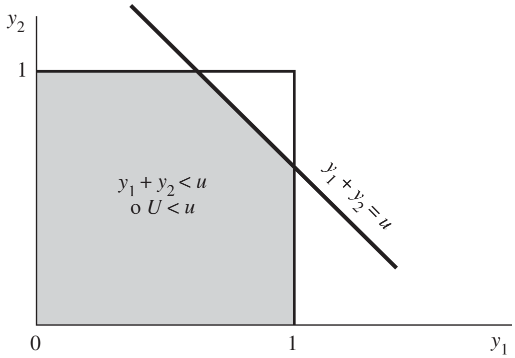


FIGURA 6.3 Región de integración para el Ejemplo 6.3

desigualdad $y_1 + y_2 < u$. Por tanto, $y_1 = y_2 = 1/4$ cae en la región sombreada debajo de la recta. Del mismo modo, todos los puntos tales que $y_1 + y_2 < u$ están debajo de la recta $y_1 + y_2 = u$. Así,

$$F_U(u) = P(U \leq u) = P(Y_1 + Y_2 \leq u) = \iint_{y_1 + y_2 \leq u} f(y_1, y_2)\, dy_1\, dy_2.$$

Si $u < 0$,

$$F_U(u) = P(U \leq u) = \iint_{y_1 + y_2 \leq u} f(y_1, y_2)\, dy_1\, dy_2 = \iint_{y_1 + y_2 \leq u} 0\, dy_1\, dy_2 = 0$$

y para $u > 2$,

$$F_U(u) = P(U \leq u) = \iint_{y_1 + y_2 \leq u} f(y_1, y_2)\, dy_1\, dy_2 = \int_0^1 \int_0^1 (1)\, dy_1\, dy_2 = 1.$$

Para $0 \leq u \leq 2$, los límites de integración dependen del valor particular de u (donde u es el punto de intersección y_2 de la recta $y_1 + y_2 = u$). En consecuencia, la expresión matemática para $F_U(u)$ cambia dependiendo de si $0 \leq u \leq 1$ o $1 < u \leq 2$.

Si $0 \leq u \leq 1$, la región $y_1 + y_2 \leq u$, es el área sombreada de la Figura 6.4. Entonces, para $0 \leq u \leq 1$, tenemos

$$F_U(u) = \iint_{y_1 + y_2 \leq u} f(y_1, y_2)\, dy_1\, dy_2 = \int_0^u \int_0^{u - y_2} (1)\, dy_1\, dy_2 = \int_0^u (u - y_2)\, dy_2$$

$$= \left(u y_2 - \frac{y_2^2}{2} \right) \Big]_0^u = u^2 - \frac{u^2}{2} = \frac{u^2}{2}.$$

La solución, $F_U(u)$, $0 \leq u \leq 1$, podría haberse obtenido directamente con geometría elemental. La densidad bivariante $f(y_1, y_2) = 1$ es uniforme en el cuadrado unitario $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$. Por tanto, $F_U(u)$ es el volumen de un sólido de altura igual a $f(y_1, y_2) = 1$ y una sección transversal triangular, como se ve en la Figura 6.4. En consecuencia,

$$F_U(u) = (\text{área de triángulo}) \cdot (\text{altura}) = \frac{u^2}{2}(1) = \frac{u^2}{2}.$$

La función de distribución se puede obtener de un modo semejante cuando u está definida en el intervalo $1 < u \leq 2$. Aun cuando la solución geométrica es más fácil, obtendremos $F_U(u)$

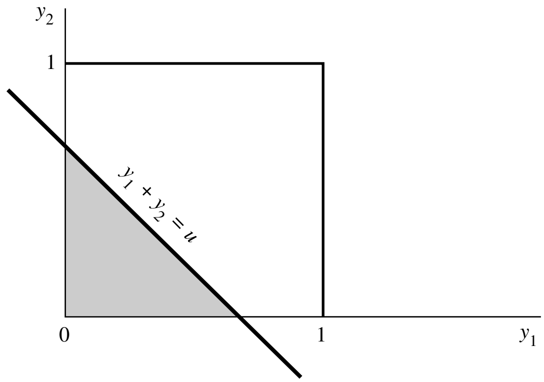


FIGURA **6.4**
Región $y_1 + y_2 \leq u$
para $0 \leq u \leq 1$

directamente por integración. La región $y_1 + y_2 \leq u$, $1 \leq u \leq 2$ es el área sombreada indicada en la Figura 6.5.

El complemento del evento $U \leq u$ es el evento de que (Y_1, Y_2) caiga en la región A de la figura 6.5. Entonces, para $1 < u \leq 2$,

$$\begin{aligned}
F_U(u) &= 1 - \int_A\int f(y_1, y_2)\,dy_1\,dy_2 \\
&= 1 - \int_{u-1}^{1}\int_{u-y_2}^{1} (1)\,dy_1\,dy_2 = 1 - \int_{u-1}^{1}\left(y_1\Big]_{u-y_2}^{1}\right)dy_2 \\
&= 1 - \int_{u-1}^{1}(1 - u + y_2)\,dy_2 = 1 - \left[(1-u)y_2 + \frac{y_2^2}{2}\right]\Bigg]_{u-1}^{1} \\
&= (-u^2/2) + 2u - 1.
\end{aligned}$$

Para resumir,

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u^2/2, & 0 \leq u \leq 1, \\ (-u^2/2) + 2u - 1, & 1 < u \leq 2, \\ 1, & u > 2. \end{cases}$$

La función de distribución para U se muestra en la Figura 6.6(a).

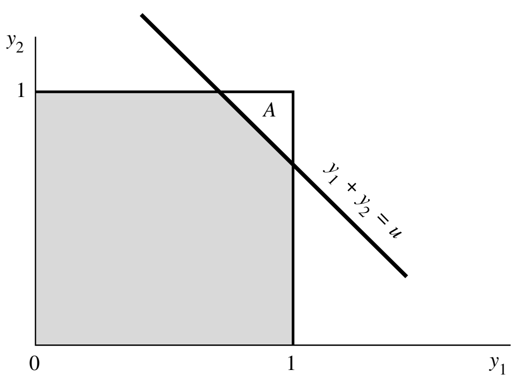
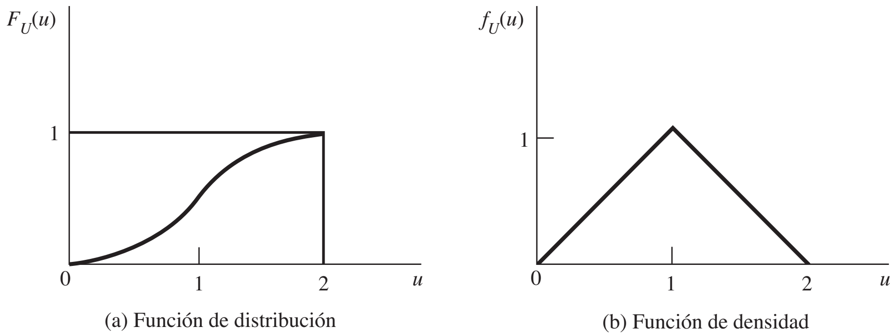


FIGURA **6.5**
Región
$y_1 + y_2 \leq u$,
$1 < u \leq 2$

FIGURA **6.6** Funciones de distribución y de densidad para el Ejemplo 6.3



La función de densidad $f_U(u)$ se puede obtener al derivar $F_U(u)$. Entonces,

$$f_U(u) = \frac{dF_U(u)}{du} \begin{cases} \frac{d}{du}(0) = 0, & u < 0, \\ \frac{d}{du}(u^2/2) = u, & 0 \le u \le 1, \\ \frac{d}{du}[(-u^2/2) + 2u - 1] = 2 - u, & 1 < u \le 2, \\ \frac{d}{du}(1) = 0, & u > 2, \end{cases}$$

o bien, lo que es más sencillo,

$$f_U(u) = \begin{cases} u, & 0 \le u \le 1, \\ 2 - u, & 1 < u \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Una gráfica de $f_U(u)$ se muestra en la Figura 6.6(b). ■

Resumen del método de las funciones de distribución

Sea U una función de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$.

1. Localice la región $U = u$ del espacio $(y_1, y_2, \ldots, y_n)$.
2. Localice la región $U \le u$.
3. Determine $F_U(u) = P(U \le u)$ al integrar $f(y_1, y_2, \ldots, y_n)$ sobre la región $U \le u$.
4. Determine la función de densidad $f_U(u)$ al derivar $F_U(u)$. Por tanto, $f_U(u) = dF_U(u)/du$.

Para ilustrar, considere el caso $U = h(Y) = Y^2$, donde Y es una variable aleatoria continua con función de distribución $F_Y(y)$ y función de densidad $f_Y(y)$. Si $u \le 0$, $F_U(u) = P(U \le u) = P(Y^2 \le u) = 0$ y para $u > 0$ (vea la Figura 6.7),

$$\begin{aligned} F_U(u) &= P(U \le u) = P(Y^2 \le u) \\ &= P(-\sqrt{u} \le Y \le \sqrt{u}) \\ &= \int_{-\sqrt{u}}^{\sqrt{u}} f(y)\,dy = F_Y(\sqrt{u}) - F_Y(-\sqrt{u}). \end{aligned}$$

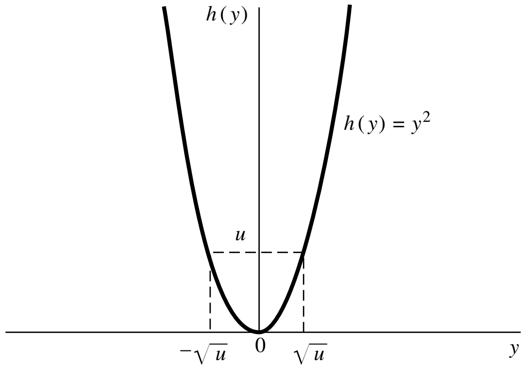


FIGURA 6.7
Función $h(y) = y^2$

En general,

$$F_U(u) = \begin{cases} F_Y(\sqrt{u}) - F_Y(-\sqrt{u}), & u > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Al derivar con respecto a u, vemos que

$$f_U(u) = \begin{cases} f_Y(\sqrt{u})\left(\dfrac{1}{2\sqrt{u}}\right) + f_Y(-\sqrt{u})\left(\dfrac{1}{2\sqrt{u}}\right), & u > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

o bien, simplemente,

$$f_U(u) = \begin{cases} \dfrac{1}{2\sqrt{u}}\left[f_Y(\sqrt{u}) + f_Y(-\sqrt{u})\right], & u > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

EJEMPLO 6.4 Sea Y una función de densidad de probabilidad dada por

$$f_Y(y) = \begin{cases} \dfrac{y+1}{2}, & -1 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad para $U = Y^2$.

Solución Sabemos que

$$f_U(u) = \begin{cases} \dfrac{1}{2\sqrt{u}}\left[f_Y(\sqrt{u}) + f_Y(-\sqrt{u})\right], & u > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y al sustituir en esta ecuación obtenemos

$$f_U(u) = \begin{cases} \dfrac{1}{2\sqrt{u}}\left(\dfrac{\sqrt{u}+1}{2} + \dfrac{-\sqrt{u}+1}{2}\right) = \dfrac{1}{2\sqrt{u}}, & 0 < u \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Como Y tiene densidad positiva sólo en el intervalo $-1 \leq y \leq 1$, se deduce que $U = Y^2$ tiene densidad positiva sólo en el intervalo $0 < u \leq 1$. ■

En algunos casos es posible encontrar una transformación que cuando se aplica a una variable aleatoria con distribución uniforme en el intervalo (0, 1), resulta en una variable aleatoria con alguna otra función de distribución específica, por ejemplo $F(y)$. El siguiente ejemplo ilustra una técnica para lograr este objetivo. A éste le sigue un breve análisis de una de las aplicaciones prácticas de esta transformación.

EJEMPLO **6.5** Sea U una variable aleatoria uniforme en el intervalo (0, 1). Encuentre una transformación $G(U)$ tal que $G(U)$ posea una distribución exponencial con media β.

Solución Si U posee una distribución uniforme en el intervalo (0, 1), entonces la función de distribución de U (vea Ejercicio 4.38) está dada por

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u, & 0 \leq u \leq 1, \\ 1, & u > 1. \end{cases}$$

Denote con Y una variable aleatoria que tiene una distribución exponencial con media β. Entonces (vea la Sección 4.6) Y tiene función de distribución

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-y/\beta}, & y \geq 0. \end{cases}$$

Observe que $F_Y(y)$ es estrictamente creciente en el intervalo $[0, \infty)$. Sea $0 < u < 1$ y advierta que hay un valor único y tal que $F_Y(y) = u$. Así, $F_Y^{-1}(u)$, $0 < u < 1$ está bien definido. En este caso, $F_Y(y) = 1 - e^{-y/\beta} = u$ si y sólo si $y = -\beta \ln(1-u) = F_Y^{-1}(u)$. Considere la variable aleatoria $F_Y^{-1}(U) = -\beta \ln(1-U)$ y observe que, si $y > 0$,

$$\begin{aligned} P\left(F_Y^{-1}(U) \leq y\right) &= P[-\beta \ln(1 - U) \leq y] \\ &= P[\ln(1 - U) \geq -y/\beta] \\ &= P(U \leq 1 - e^{-y/\beta}) \\ &= 1 - e^{-y/\beta}. \end{aligned}$$

También, $P\left[F_Y^{-1}(U) \leq y\right] = 0$ si $y \leq 0$. Así, $F_Y^{-1}(U) = -\beta \ln(1 - U)$ posee una distribución exponencial con media β, como se desea. ■

Las simulaciones por computadora se usan con frecuencia para evaluar técnicas estadísticas propuestas. Por lo general estas simulaciones requieren que obtengamos valores observados de variables aleatorias con distribución prescrita. Como se observó en la Sección 4.4, casi todos los sistemas computarizados contienen una subrutina que genera valores observados de una variable aleatoria U que tiene una distribución uniforme en el intervalo (0, 1). ¿Cómo puede usarse el resultado del Ejemplo 6.5 para generar un conjunto de observaciones a partir de una distribución exponencial con media β? Simplemente usamos el generador de números aleatorio de la computadora para producir valores $u_1, u_2, \ldots, u_n$ de una distribución uniforme (0, 1) y luego calculamos $y_i = -\beta \ln(1 - u_i)$, $i = 1, 2, \ldots, n$ para obtener valores de variables aleatorias con la distribución exponencial requerida.

Mientras una función de distribución prescrita $F(y)$ posea una inversa única $F^{-1}(\cdot)$, se puede aplicar la técnica anterior. En casos como el ilustrado en el Ejemplo 6.5 podemos fácilmente escribir la forma de $F^{-1}(\cdot)$ y continuar como antes. Si la expresión de una función de distribución no se puede escribir en una forma fácilmente invertible (recuerde que las funciones de distribución de variables aleatorias de distribución normal, gamma y beta se determinan por medio de tablas que se obtuvieron con el uso de técnicas de integración numérica), nuestro trabajo es más difícil. En estos casos se usan otros métodos para generar observaciones con la distribución deseada.

En el siguiente conjunto de ejercicios encontrará problemas que se pueden resolver con el uso de las técnicas presentadas en esta sección. Los ejercicios que implican determinar $F^{-1}(U)$ para alguna distribución específica $F(y)$ se concentran en casos donde $F^{-1}(\cdot)$ existe en forma cerrada.

Ejercicios

6.1 Sea Y una variable aleatoria con función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 2(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la función de densidad de $U_1 = 2Y - 1$.
b Encuentre la función de densidad de $U_2 = 1 - 2Y$.
c Encuentre la función de densidad de $U_3 = Y^2$.
d Encuentre $E(U_1)$, $E(U_2)$ y $E(U_3)$ usando las funciones de densidad obtenidas para estas variables aleatorias.
e Encuentre $E(U_1)$, $E(U_2)$ y $E(U_3)$ con los métodos del Capítulo 4.

6.2 Sea Y una variable aleatoria con una función de densidad dada por

$$f(y) = \begin{cases} (3/2)y^2, & -1 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la función de densidad de $U_1 = 3Y$.
b Encuentre la función de densidad de $U_2 = 3 - Y$.
c Encuentre la función de densidad de $U_3 = Y^2$.

6.3 Un proveedor de queroseno tiene una demanda semanal Y que posee una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} y, & 0 \le y \le 1, \\ 1, & 1 < y \le 1.5, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

con mediciones en cientos de galones. (Este problema planteó en el Ejercicio 4.13.) La utilidad del proveedor está dada por $U = 10Y - 4$.

a Encuentre la función de densidad de probabilidad para U.
b Use la respuesta al inciso a para encontrar $E(U)$.
c Encuentre $E(U)$ con los métodos del Capítulo 4.

6.4 La cantidad de harina consumida por día en una panadería es una variable aleatoria Y que tiene una distribución exponencial con media igual a 4 toneladas. El costo de la harina es proporcional a $U = 3Y + 1$.

a Encuentre la función de densidad de probabilidad para U.

b Use la respuesta del inciso a para hallar $E(U)$.

6.5 El tiempo de espera Y hasta la entrega de un nuevo componente para una operación industrial está uniformemente distribuido en el intervalo de 1 a 5 días. El costo de esta demora está dado por $U = 2Y^2 + 3$. Encuentre la función de densidad de probabilidad para U.

6.6 La distribución conjunta de la cantidad de contaminantes emitida desde una chimenea sin aparato limpiador (Y_1) y una chimenea igual pero con aparato limpiador (Y_2) se dio en el Ejercicio 5.10 como

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 2,\ 0 \le y_2 \le 1,\ 2y_2 \le y_1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La reducción en la cantidad de contaminantes debida al aparato limpiador está dada por $U = Y_1 - Y_2$.

a Encuentre la función de densidad de probabilidad para U.

b Use la respuesta del inciso a para hallar $E(U)$. Compare sus resultados con los del Ejercicio 5.78(c).

6.7 Suponga que Z tiene una distribución normal estándar.

a Encuentre la función de densidad de $U = Z^2$.

b ¿U tiene una distribución gamma? ¿Cuáles son los valores de α y β?

c ¿Cuál es otro nombre para la distribución de U?

6.8 Suponga que Y tiene una distribución beta con parámetros α y β.

a Encuentre la función de densidad de $U = 1 - Y$.

b Identifique la densidad de U como uno de los tipos que estudiamos en el Capítulo 4. Asegúrese de identificar cualesquiera valores de parámetro.

c ¿Cómo están relacionadas $E(U)$ y $E(Y)$?

d ¿Cómo están relacionadas $V(U)$ y $V(Y)$?

6.9 Suponga que una unidad de mineral contiene una proporción Y_1 de metal A y una proporción Y_2 de metal B. La experiencia ha demostrado que la función de densidad de probabilidad conjunta de Y_1 y Y_2 es uniforme en la región $0 \le y_1 \le 1$, $0 \le y_2 \le 1$, $0 \le y_1 + y_2 \le 1$. Sea $U = Y_1 + Y_2$, la proporción del metal A o B por unidad. Encuentre

a la función de densidad de probabilidad para U.

b $E(U)$ usando la respuesta al inciso a.

c $E(U)$ usando sólo las densidades marginales de Y_1 y Y_2.

6.10 El tiempo total desde la llegada hasta la terminación del servicio en un restaurante de comida rápida, Y_1, y el tiempo empleado en la fila de espera antes de llegar a la ventanilla de servicio, Y_2, se dieron en el Ejercicio 5.15 con función de densidad conjunta

$$f(y_1, y_2) = \begin{cases} e^{-y_1}, & 0 \le y_2 \le y_1 < \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Otra variable aleatoria de interés es $U = Y_1 - Y_2$, el tiempo empleado en la ventanilla de servicio. Encuentre

a la función de densidad de probabilidad para U.

b $E(U)$ y $V(U)$. Compare sus respuestas con los resultados del Ejercicio 5.108.

6.11 Suponga que dos componentes electrónicos del sistema de guía de un proyectil operan de manera independiente y que cada uno tiene una vida útil gobernada por la distribución exponencial con media 1 (con mediciones en cientos de horas). Encuentre la

a función de densidad de probabilidad para el promedio de vida útil de los dos componentes.

b media y la varianza de este promedio, usando la respuesta del inciso a. Compruebe su respuesta al calcular la media y la varianza con el Teorema 5.12.

6.12 Suponga que Y tiene una distribución gamma con parámetros α y β y que $c > 0$ es una constante.

a Deduzca la función de densidad de $U = cY$.

b Identifique la densidad de U como uno de los tipos que estudiamos en el Capítulo 4. Asegúrese de identificar cualesquiera valores de parámetro.

c Los parámetros α y β de una variable aleatoria con distribución gamma son, respectivamente, parámetros de "forma" y "escala". ¿Cómo se comparan los parámetros de escala y forma para U con los de Y?

6.13 Si Y_1 y Y_2 son variables aleatorias exponenciales independientes, ambas con media β, encuentre la función de densidad para la suma de ellas. (En el Ejercicio 5.7 consideramos dos variables aleatorias exponenciales independientes, ambas con media 1 y $P(Y_1 + Y_2 \leq 3)$ determinada.)

6.14 En un proceso de sinterización (calentamiento) de dos tipos de polvo de cobre (vea el Ejercicio 5.152), la función de densidad para Y_1, la proporción de volumen de cobre sólido en una muestra, estuvo dada por

$$f_1(y_1) = \begin{cases} 6y_1(1 - y_1), & 0 \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La función de densidad para Y_2, la proporción de cristales tipo A entre el cobre sólido, estuvo dada por

$$f_2(y_2) = \begin{cases} 3y_2^2, & 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La variable $U = Y_1Y_2$ da la proporción del volumen de la muestra debida a los cristales tipo A. Si Y_1 y Y_2 son independientes, encuentre la función de densidad de probabilidad para U.

6.15 Sea Y una función de distribución dada por

$$F(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-y^2}, & y \geq 0. \end{cases}$$

Encuentre una transformación $G(U)$ tal que, si U tiene una distribución uniforme en el intervalo (0, 1), $G(U)$ tiene la misma distribución que Y.

6.16 En el Ejercicio 4.15 determinamos que

$$f(y) = \begin{cases} \dfrac{b}{y^2}, & y \geq b, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

es una función de densidad de probabilidad auténtica para una variable aleatoria, Y. Suponiendo que b es una constante conocida y U tiene una distribución uniforme en el intervalo (0, 1), transforme U para obtener una variable aleatoria con la misma distribución que Y.

6.17 Un miembro de la familia de distribuciones de potencia tiene una función de distribución dada por

$$F(y) = \begin{cases} 0, & y < 0, \\ \left(\dfrac{y}{\theta}\right)^{\alpha}, & 0 \leq y \leq \theta, \\ 1, & y > \theta, \end{cases}$$

donde $\alpha, \theta > 0$.

a Encuentre la función de densidad.

b Para valores fijos de α y θ encuentre una transformación $G(U)$ de modo que $G(U)$ tenga una función de distribución de F cuando U posea una distribución uniforme (0, 1).

c Dado que una muestra aleatoria de tamaño 5 tomada de una distribución uniforme en el intervalo (0, 1) dio los valores .2700, .6901, .1413, .1523 y .3609, use la transformación que obtuvo en el inciso b para dar valores asociados con una variable aleatoria con una familia de distribución de potencia con $\alpha = 2, \theta = 4$.

6.18 Un miembro de la familia de distribuciones de Pareto (que se usa a menudo en economía para modelar distribuciones de ingreso) tiene una función de distribución dada por

$$F(y) = \begin{cases} 0, & y < \beta, \\ 1 - \left(\dfrac{\beta}{y}\right)^{\alpha}, & y \geq \beta, \end{cases}$$

donde $\alpha, \beta > 0$.

a Encuentre la función de densidad.

b Para valores fijos de β y α, encuentre una transformación $G(U)$ de modo que $G(U)$ tenga una función de distribución de F cuando U posea una distribución uniforme en el intervalo (0, 1).

c Dado que una muestra aleatoria de tamaño 5 tomada de una distribución uniforme en el intervalo (0, 1) dio los valores .0058, .2048, .7692, .2475 y .6078, use la transformación que dedujo en el inciso b para dar valores asociados con una variable aleatoria con una distribución de Pareto con $\alpha = 2$, $\beta = 3$.

6.19 Consulte los Ejercicios 6.17 y 6.18. Si Y posee una distribución de Pareto con parámetros α y β, demuestre que $X = 1/Y$ tiene una familia de distribución de potencia con parámetros α y $\theta = \beta^{-1}$.

6.20 Sea la variable aleatoria Y que tiene una distribución uniforme en el intervalo (0, 1). Deduzca

a la distribución de la variable aleatoria $W = Y^2$,

b la distribución de la variable aleatoria $W = \sqrt{Y}$.

***6.21** Suponga que Y es una variable aleatoria que toma sólo valores enteros 1, 2, . . . Denote con $F(y)$ la función de distribución de esta variable aleatoria. Como ya vimos en la Sección 4.2, esta función de distribución es una función escalón y la magnitud del escalón en cada valor entero es la probabilidad de que Y tome ese valor. Sea U una variable aleatoria continua que está uniformemente distribuida en el intervalo (0, 1). Defina una variable X tal que $X = k$ si y sólo si $F(k - 1) < U \leq F(k)$, $k = 1, 2, \ldots$ Recuerde que $F(0) = 0$ porque Y toma sólo valores enteros *positivos*. Demuestre que $P(X = i) = F(i) - F(i - 1) = P(Y = i)$, $i = 1, 2, \ldots$ Esto es, X tiene la misma distribución que Y. [*Sugerencia:* recuerde el Ejercicio 4.5.][1]

***6.22** Use los resultados que obtuvo en los Ejercicios 4.6 y 6.21 para describir cómo generar valores de una variable aleatoria geométricamente distribuida.

6.4 Método de las transformaciones

El método de las transformaciones, el cual nos permite determinar la distribución de probabilidad de una función de variables aleatorias, es una consecuencia del método de función de distribución de la Sección 6.3. Mediante el método de las funciones de distribución podemos llegar a un método simple para formular la función de densidad de $U = h(Y)$, siempre que $h(y)$

1. Los ejercicios precedidos de un asterisco son opcionales.

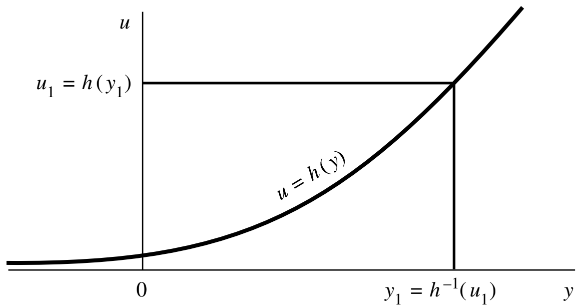


FIGURA **6.8** Una función creciente

sea decreciente o creciente. [Por $h(y)$ creciente, queremos decir que si $y_1 < y_2$, entonces $h(y_1) < h(y_2)$ para cualesquiera números reales y_1 y y_2.] La gráfica de una función creciente $h(y)$ aparece en la Figura 6.8.

Suponga que $h(y)$ es una función creciente de y y que $U = h(Y)$, donde Y tiene función de densidad $f_Y(y)$. Entonces $h^{-1}(u)$ es una función creciente de u: si $u_1 < u_2$, entonces $h^{-1}(u_1) = y_1 < y_2 = h^{-1}(u_2)$. En la figura 6.8 vemos que el conjunto de puntos y tales que $h(y) \le u_1$ es precisamente igual que el conjunto de puntos y tales que $y \le h^{-1}(u_1)$. Por tanto (vea la Figura 6.8),

$$P(U \le u) = P[h(Y) \le u] = P\{h^{-1}[h(Y)] \le h^{-1}(u)\} = P[Y \le h^{-1}(u)]$$

o bien,

$$F_U(u) = F_Y[h^{-1}(u)].$$

Entonces derivando con respecto a u, tenemos

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y[h^{-1}(u)]}{du} = f_Y(h^{-1}(u))\frac{d[h^{-1}(u)]}{du}.$$

Para simplificar la notación, escribiremos dh^{-1}/du en lugar de $d[h^{-1}(u)]/du$ y

$$f_U(u) = f_Y[h^{-1}(u)]\frac{dh^{-1}}{du}.$$

En consecuencia, tenemos una nueva forma para determinar $f_U(u)$ que surge a partir del método general de funciones de distribución. Para determinar $f_U(u)$, despeje y en términos de u; esto es, encuentre $y = h^{-1}(u)$ y sustituya esta expresión en $f_Y(y)$. Entonces multiplique esta cantidad por dh^{-1}/du. Ilustraremos el procedimiento con un ejemplo.

EJEMPLO **6.6** En el Ejemplo 6.1 trabajamos con una variable aleatoria Y (cantidad de azúcar producida) con una función de densidad dada por

$$f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Estamos interesados en una nueva variable aleatoria (utilidad) dada por $U = 3Y - 1$. Encuentre la función de densidad de probabilidad para U por el método de las transformaciones.

Solución La función de interés aquí es $h(y) = 3y - 1$, que es creciente en y. Si $u = 3y - 1$, entonces

$$y = h^{-1}(u) = \frac{u+1}{3} \quad \text{y} \quad \frac{dh^{-1}}{du} = \frac{d\left(\frac{u+1}{3}\right)}{du} = \frac{1}{3}.$$

Por tanto,

$$\begin{aligned} f_U(u) &= f_Y[h^{-1}(u)]\frac{dh^{-1}}{du} \\ &= \begin{cases} 2[h^{-1}(u)]\dfrac{dh^{-1}}{du} = 2\left(\dfrac{u+1}{3}\right)\left(\dfrac{1}{3}\right), & 0 \le \dfrac{u+1}{3} \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases} \end{aligned}$$

o bien, lo que es equivalente,

$$f_U(u) = \begin{cases} 2(u+1)/9, & -1 \le u \le 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El intervalo en el cual $f_U(u)$ es positiva es simplemente $0 \le y \le 1$ transformado en el eje u por la función $u = 3y - 1$. Esta respuesta está acorde con la del Ejemplo 6.1. ■

Si $h(y)$ es una función decreciente de y, entonces $h^{-1}(u)$ es una función decreciente de u. Esto es, si $u_1 < u_2$, entonces $h^{-1}(u) = y_1 > y_2 = h^{-1}(u_2)$. También, como en la Figura 6.9, el conjunto de puntos y tales que $h(y) \le u_1$ es el mismo conjunto de puntos tales que $y \ge h^{-1}(u_1)$.

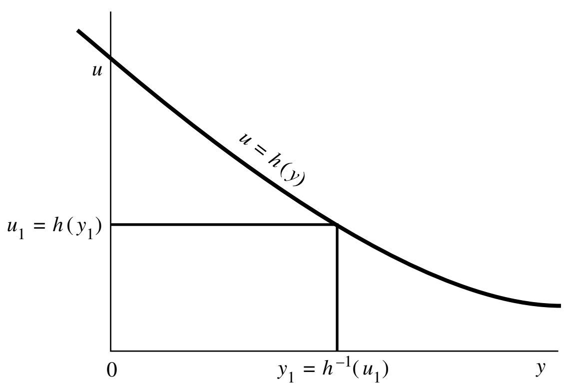
Se deduce que, para $U = h(Y)$, como se ve en la Figura 6.9,

$$P(U \le u) = P[Y \ge h^{-1}(u)] \quad \text{o} \quad F_U(u) = 1 - F_Y[h^{-1}(u)].$$

Si derivamos con respecto a u, obtenemos

$$f_U(u) = -f_Y[h^{-1}(u)]\frac{d[h^{-1}(u)]}{du}.$$

FIGURA **6.9** Una función decreciente



Si de nuevo empleamos la notación simplificada dh^{-1}/du en lugar de $d\,[h^{-1}(u)]/du$ y recordamos que dh^{-1}/du es negativa porque $h^{-1}(u)$ es una función decreciente de u, la densidad de U es

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|.$$

En realidad, no es necesario que $h(y)$ sea creciente o decreciente (y por tanto que se pueda invertir) para todos los valores de y. La función $h(\cdot)$ sólo necesita ser creciente o decreciente para los valores de y tales que $f_Y(y) > 0$. El conjunto de puntos $\{y: f_Y(y) > 0\}$ se denomina *soporte* de la densidad $f_Y(y)$. Si $y = h^{-1}(u)$ no es el soporte de la densidad, entonces $f_Y[h^{-1}(u)] = 0$. Estos resultados se combinan en el siguiente enunciado:

Sea Y la función de densidad de probabilidad $f_Y(y)$. Si $h(y)$ es creciente o decreciente para toda y tal que $f_Y(y) > 0$, entonces $U = h(Y)$ tiene función de densidad

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|, \qquad \text{donde} \quad \frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}.$$

EJEMPLO **6.7** Sea Y la función de densidad de probabilidad dada por

$$f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad de $U = -4Y + 3$.

Solución En este ejemplo, el conjunto de valores de y tales que $f_Y(y) > 0$ son los valores $0 < y \le 1$. La función de interés, $h(y) = -4y + 3$, es decreciente para toda y y, por tanto, para toda $0 < y \le 1$, si $u = -4y + 3$, entonces

$$y = h^{-1}(u) = \frac{3-u}{4} \quad \text{y} \quad \frac{dh^{-1}}{du} = -\frac{1}{4}.$$

Observe que $h^{-1}(u)$ es una función decreciente de u y que $dh^{-1}/du < 0$. Entonces,

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right| = \begin{cases} 2\left(\dfrac{3-u}{4}\right)\left|-\dfrac{1}{4}\right|, & 0 \le \dfrac{3-u}{4} \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Finalmente, un poco de álgebra nos da

$$f_U(u) = \begin{cases} \dfrac{3-u}{8}, & -1 \le u \le 3, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

■

La aplicación directa del método de las transformaciones requiere que la función $h(y)$ sea creciente o decreciente para toda y tal que $f_Y(y) > 0$. Si desea usar este método para hallar la distribución de $U = h(Y)$, debe tener mucho cuidado de comprobar que la función $h(\cdot)$ sea creciente o decreciente para toda y en el soporte de $f_Y(y)$. Si no es así, el método de las transformaciones no se puede aplicar y en su lugar se puede utilizar el método de las funciones de distribución estudiadas en la Sección 6.3.

El método de las transformaciones también se puede emplear en situaciones multivariantes. El siguiente ejemplo ilustra el caso bivariante.

EJEMPLO **6.8** Sean Y_1 y Y_2 que tienen una función de densidad conjunta dada por

$$f(y_1, y_2) = \begin{cases} e^{-(y_1+y_2)}, & 0 \le y_1,\ 0 \le y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad para $U = Y_1 + Y_2$.

Solución Este problema se debe resolver en dos etapas: primero, determinaremos la densidad conjunta de Y_1 y U y en seguida la densidad marginal de U. El método consiste en asignar a Y_1 un valor $y_1 \ge 0$. Entonces $U = y_1 + Y_2$ y podemos considerar el problema de transformación unidimensional en el que $U = h(Y_2) = y_1 + Y_2$. Si $g(y_1, u)$ denota la densidad conjunta de Y_1 y U, tenemos, con $y_2 = u - y_1 = h^{-1}(u)$,

$$g(y_1, u) = \begin{cases} f[y_1, h^{-1}(u)]\left|\dfrac{dh^{-1}}{du}\right| = e^{-(y_1+u-y_1)}(1), & 0 \le y_1,\ 0 \le u - y_1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Simplificando obtendremos

$$g(y_1, u) = \begin{cases} e^{-u}, & 0 \le y_1 \le u, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(Observe que $Y_1 \le U$). La densidad marginal de U está dada entonces por

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} g(y_1, u)\, dy_1 \\ &= \begin{cases} \displaystyle\int_0^u e^{-u}\, dy_1 = ue^{-u}, & 0 \le u, \\ 0, & \text{en cualquier otro punto.} \end{cases} \end{aligned}$$

■

Ilustraremos el uso de la transformación bivariante con otro ejemplo que implique el producto de dos variables aleatorias.

EJEMPLO **6.9** En el Ejemplo 5.19 consideramos una variable aleatoria Y_1 como la proporción de impurezas en una muestra química y Y_2 como la proporción de impurezas tipo I entre todas las impurezas de la muestra. La función de densidad conjunta estuvo dada por

$$f(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \le y_1 \le 1,\ 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Estamos interesados en $U = Y_1Y_2$, que es la proporción de impurezas tipo I de la muestra. Encuentre la función de densidad de probabilidad para U y utilícela para hallar $E(U)$.

Solución Como estamos interesados en $U = Y_1Y_2$, primero asignamos a Y_1 un valor y_1, $0 < y_1 \leq 1$ y pensamos en términos de la transformación univariante $U = h(Y_2) = y_1Y_2$. Podemos entonces determinar la función de densidad conjunta para Y_1 y U (con $y_2 = u/y_1 = h^{-1}(u)$ como

$$g(y_1, u) = f[y_1, h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|$$
$$= \begin{cases} 2(1 - y_1)\left|\dfrac{1}{y_1}\right|, & 0 < y_1 \leq 1,\ 0 \leq u/y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

De modo equivalente, se puede expresar como

$$g(y_1, u) = \begin{cases} 2(1 - y_1)\left(\dfrac{1}{y_1}\right), & 0 \leq u \leq y_1 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(U también varía entre 0 y 1, pero Y_1 siempre debe ser mayor o igual a U.) Además,

$$f_U(u) = \int_{-\infty}^{\infty} g(y_1, u)\, dy_1$$
$$= \begin{cases} \displaystyle\int_u^1 2(1 - y_1)\left(\frac{1}{y_1}\right) dy_1, & 0 \leq u \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Para $0 \leq u \leq 1$,

$$\int_u^1 2(1 - y_1)\left(\frac{1}{y_1}\right) dy_1 = 2\int_u^1 \left(\frac{1}{y_1} - 1\right) dy_1$$
$$= 2\left(\ln y_1\Big]_u^1 - y_1\Big]_u^1\right) = 2\,(-\ln u - 1 + u)$$
$$= 2(u - \ln u - 1),$$

obtenemos

$$f_U(u) = \begin{cases} 2(u - \ln u - 1), & 0 \leq u \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(El símbolo ln significa logaritmo natural.)

Ahora encontramos $E(U)$:

$$E(U) = \int_{-\infty}^{\infty} u f_U(u)\, du = \int_0^1 2u(u - \ln u - 1)\, du$$
$$= 2\left\{\int_0^1 u^2\, du - \int_0^1 u(\ln u)\, du - \int_0^1 u\, du\right\}$$
$$= 2\left\{\frac{u^3}{3}\Big]_0^1 - \int_0^1 u(\ln u)\, du - \frac{u^2}{2}\Big]_0^1\right\}.$$

La integral de en medio se resuelve con más facilidad si se usa integración por partes, lo cual dará

$$\int_0^1 u(\ln u)\,du = \left(\frac{u^2}{2}\right)(\ln u)\Big]_0^1 - \int_0^1 \left(\frac{u^2}{2}\right)\left(\frac{1}{u}\right)du = 0 - \frac{u^2}{4}\Big]_0^1 = -\frac{1}{4}.$$

Entonces,

$$E(U) = 2[(1/3) - (-1/4) - (1/2)] = 2(1/12) = 1/6.$$

Esta respuesta concuerda con la del Ejemplo 5.21, donde $E(U) = E(Y_1Y_2)$ se dedujo con un método diferente. ■

Resumen del método de las transformaciones

Sea $U = h(Y)$, donde $h(y)$ es una función creciente o decreciente de y para toda y tal que $f_Y(y) > 0$.

1. Deduzca la función inversa, $y = h^{-1}(u)$.
2. Evalúe $\dfrac{dh^{-1}}{du} = \dfrac{d[h^{-1}(u)]}{du}$.
3. Encuentre $f_U(u)$ mediante

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|.$$

Ejercicios

6.23 En el Ejercicio 6.1 consideramos una variable aleatoria Y con función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 2(1-y), & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y empleamos el método de las funciones de distribución para determinar las funciones de densidad de

a $U_1 = 2Y - 1$.

b $U_2 = 1 - 2Y$.

c $U_3 = Y^2$.

Use el método de las transformaciones para hallar las densidades de U_1, U_2 y U_3.

6.24 En el Ejercicio 6.4, consideramos una variable aleatoria Y que poseía una distribución exponencial con media 4 y usamos el método de las funciones de distribución para obtener la función de densidad para $U = 3Y + 1$. Utilice el método de las transformaciones para deducir la función de densidad para U.

6.25 En el Ejercicio 6.11 consideramos dos componentes electrónicos que operan de modo independiente, cada uno con vida útil gobernada por la distribución exponencial con media 1. Procedimos a usar el método de las funciones de distribución para obtener la distribución del promedio de vida útil para los dos componentes. Utilice el método de las transformaciones para obtener la función de densidad para el promedio de vida útil de los dos componentes.

6.26 La función de densidad de Weibull está dada por

$$f(y) = \begin{cases} \dfrac{1}{\alpha} m y^{m-1} e^{-y^m/\alpha}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde α y m son constantes positivas. Esta función de densidad se usa con frecuencia como modelo para la vida útil de sistemas físicos. Suponga que Y tiene la densidad de Weibull dada. Encuentre

a la función de densidad de $U = Y^m$.

b $E(Y^k)$ para cualquier entero positivo k.

6.27 Sea Y que tiene una distribución exponencial con media β.

a Demuestre que $W = \sqrt{Y}$ tiene una densidad de Weibull con $\alpha = \beta$ y $m = 2$.

b Use el resultado del Ejercicio 6.26(b) para deducir $E(Y^{k/2})$ para cualquier entero positivo k.

6.28 Sea Y con una distribución uniforme (0, 1). Demuestre que $U = -2\ln(Y)$ tiene una distribución exponencial con media 2.

6.29 La velocidad de una molécula en un gas uniforme en equilibrio es una variable aleatoria V cuya función de densidad está dada por

$$f(v) = av^2 e^{-bv^2}, \quad v > 0,$$

donde $b = m/2kT$ y k, T y m denotan la constante de Boltzmann, la temperatura absoluta y la masa de la molécula, respectivamente.

a Deduzca la distribución de $W = mV^2/2$, la energía cinética de la molécula.

b Encuentre $E(W)$.

6.30 Una corriente eléctrica fluctuante I puede ser considerada una variable aleatoria uniformemente distribuida en el intervalo (9, 11). Si esta corriente pasa por un resistor de 2 ohms, encuentre la función de densidad de probabilidad de la potencia $P = 2I^2$.

6.31 La distribución conjunta para la vida útil de dos diferentes tipos de componentes que operan en un sistema se determinó en el Ejercicio 5.18 por

$$f(y_1, y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0,\ y_2 > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La eficiencia relativa de los dos tipos de componentes está medida por $U = Y_2/Y_1$. Encuentre la función de densidad de probabilidad para U.

6.32 En el Ejercicio 6.5 consideramos una variable aleatoria Y que tiene una distribución uniforme en el intervalo [1, 5]. El costo de demora está dado por $U = 2Y^2 + 3$. Use el método de las transformaciones para obtener la función de densidad de U.

6.33 La proporción de impurezas en ciertas muestras de mineral es una variable aleatoria Y con función de densidad dada por

$$f(y) = \begin{cases} (3/2)y^2 + y, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El valor en dólares de esas muestras es $U = 5 - (Y/2)$. Encuentre la función de densidad de probabilidad para U.

6.34 Una función de densidad que a veces utilizan ingenieros para modelar duraciones de vida útil de componentes electrónicos es la densidad de Rayleigh, dada por

$$f(y) = \begin{cases} \left(\dfrac{2y}{\theta}\right)e^{-y^2/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Si Y tiene la densidad de Rayleigh, encuentre la función de densidad de probabilidad para $U = Y^2$.

b Utilice el resultado del inciso a para hallar $E(Y)$ y $V(Y)$.

6.35 Sean Y_1 y Y_2 variables aleatorias independientes, ambas uniformemente distribuidas en (0, 1). Encuentre la función de densidad de probabilidad para $U = Y_1Y_2$.

6.36 Consulte el Ejercicio 6.34. Sean Y_1 y Y_2 variables aleatorias independientes con distribución de Rayleigh. Encuentre la función de densidad de probabilidad para $U = Y_1^2 + Y_2^2$. [*Sugerencia:* recuerde el Ejemplo 6.8.]

6.5 Método de las funciones generadoras de momento

El método de las funciones generadoras de momento para determinar la distribución de probabilidad de una función de variables aleatorias $Y_1, Y_2, \ldots, Y_n$ está basada en el siguiente teorema de unicidad.

TEOREMA 6.1 Denotemos con $m_X(t)$ y $m_Y(t)$ las funciones generadoras de momento de variables aleatorias X y Y, respectivamente. Si existen funciones generadoras de momento y $m_X(t) = m_Y(t)$ para todos los valores de t, entonces X y Y tienen la misma distribución de probabilidad.

(La demostración del Teorema 6.1 está fuera del alcance de este libro.)

Si U es una función de n variables aleatorias $Y_1, Y_2, \ldots, Y_n$, el primer paso al usar el Teorema 6.1 es hallar la función generadora de momento de U:

$$m_U(t) = E(e^{tU}).$$

Una vez determinada la función generadora de momento para U, se compara con las funciones generadoras de momento para variables aleatorias con distribuciones bien conocidas. Si $m_U(t)$ es idéntica a una de éstas, por ejemplo la función generadora de momento para una variable aleatoria V, entonces, por el Teorema 6.1, U y V poseen distribuciones de probabilidad idénticas. Las funciones de densidad, medias, varianzas y funciones generadoras de momento para algunas variables aleatorias que se encuentran con frecuencia se presentan en el Apéndice 2. Ilustraremos el procedimiento con unos pocos ejemplos.

EJEMPLO 6.10 Sea Y una variable aleatoria normalmente distribuida con media μ y varianza σ^2. Demuestre que

$$Z = \frac{Y - \mu}{\sigma}$$

tiene una distribución *normal estándar*, una distribución normal con media 0 y varianza 1.

Solución Hemos visto en el Ejemplo 4.16 que $Y - \mu$ tiene función generadora de momento $e^{t^2\sigma^2/2}$. Por tanto,

$$m_Z(t) = E(e^{tZ}) = E[e^{(t/\sigma)(Y-\mu)}] = m_{(Y-\mu)}\left(\frac{t}{\sigma}\right) = e^{(t/\sigma)^2(\sigma^2/2)} = e^{t^2/2}.$$

Al comparar $m_Z(t)$ con la función generadora de momento de una variable aleatoria normal, vemos que Z debe estar normalmente distribuida con $E(Z) = 0$ y $V(Z) = 1$. ■

EJEMPLO **6.11** Sea Z una variable aleatoria normalmente distribuida con media 0 y varianza 1. Use el método de las funciones generadoras de momento para determinar la distribución de probabilidad de Z^2.

Solución La función generadora de momentos para Z^2 es

$$m_{Z^2}(t) = E(e^{tZ^2}) = \int_{-\infty}^{\infty} e^{tz^2} f(z)\, dz = \int_{-\infty}^{\infty} e^{tz^2} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)(1-2t)}\, dz.$$

Esta integral se puede evaluar ya sea consultando una tabla de integrales o tomando en cuenta que, si $1 - 2t > 0$ (de manera equivalente, $t < 1/2$), el integrando

$$\frac{\exp\left[-\left(\frac{z^2}{2}\right)(1-2t)\right]}{\sqrt{2\pi}} = \frac{\exp\left[-\left(\frac{z^2}{2}\right)\Big/(1-2t)^{-1}\right]}{\sqrt{2\pi}}$$

es proporcional a la función de densidad de una variable aleatoria normalmente distribuida con media 0 y varianza $(1 - 2t)^{-1}$. Para hacer del integrando una función de densidad normal (para que la integral definida sea igual a 1), multiplicamos el numerador y denominador por la desviación estándar, $(1 - 2t)^{-1/2}$. Entonces

$$m_{Z^2}(t) = \frac{1}{(1-2t)^{1/2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-1/2}} \exp\left[-\left(\frac{z^2}{2}\right)\Big/(1-2t)^{-1}\right] dz.$$

Como la integral es igual a 1, si $t < 1/2$,

$$m_{Z^2}(t) = \frac{1}{(1-2t)^{1/2}} = (1-2t)^{-1/2}.$$

Una comparación de $m_{Z^2}(t)$ con las funciones generadoras de momento del Apéndice 2 muestra que $m_{Z^2}(t)$ es idéntica a la función generadora de momento para la variable aleatoria con distribución gamma y $\alpha = 1/2$ y $\beta = 2$. Así, usando la Definición 4.10, Z^2 tiene una distribución χ^2 con $\nu = 1$ grado de libertad. Se deduce que la función de densidad para $U = Z^2$ está dada por

$$f_U(u) = \begin{cases} \dfrac{u^{-1/2} e^{-u/2}}{\Gamma(1/2) 2^{1/2}}, & u \geq 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$ ■

El método de las funciones generadoras de momento es con frecuencia muy útil para hallar las distribuciones de sumas de variables aleatorias independientes.

TEOREMA 6.2 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes con funciones generadoras de momento $m_{Y_1}(t), m_{Y_2}(t), \ldots, m_{Y_n}(t)$, respectivamente. Si $U = Y_1 + Y_2 + \cdots + Y_n$, entonces

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t).$$

Demostración Sabemos que, como las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ son independientes (vea el Teorema 5.9),

$$\begin{aligned} m_U(t) &= E\left[e^{t(Y_1+\cdots+Y_n)}\right] = E\left(e^{tY_1}e^{tY_2}\cdots e^{tY_n}\right) \\ &= E\left(e^{tY_1}\right) \times E\left(e^{tY_2}\right) \times \cdots \times E\left(e^{tY_n}\right). \end{aligned}$$

Entonces, por la definición de funciones generadoras de momentos,

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t).$$

EJEMPLO 6.12 La cantidad de clientes que llegan a una caja para pagar en un intervalo determinado de tiempo posee aproximadamente una distribución de probabilidad de Poisson (vea la Sección 3.8). Si Y_1 denota el tiempo que transcurre hasta la primera llegada, Y_2 denota el tiempo entre la primera y la segunda llegadas, . . . , y Y_n denota el tiempo entre la llegada del primer $(n - 1)$ cliente y la n-ésima llegada, entonces se puede demostrar que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes, con la función de densidad para Y_i dada por

$$f_{Y_i}(y_i) = \begin{cases} \dfrac{1}{\theta} e^{-y_i/\theta}, & y_i > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

[Como las variables Y_i, para $i = 1, 2, \ldots, n$, están distribuidas exponencialmente, se deduce que $E(Y_i) = \theta$; esto es, θ es el tiempo promedio que transcurre entre llegadas.] Encuentre la función de densidad de probabilidad para el tiempo de espera desde que se abre la caja de salida hasta que llega el n-ésimo cliente. (Si $Y_1, Y_2, \ldots$ denotan tiempos sucesivos entre llegadas, buscamos la función de densidad de $U = Y_1 + Y_2 + \cdots + Y_n$.)

Solución Para aplicar el Teorema 6.2 debemos conocer primero $m_{Y_i}(t)$, $i = 1, 2, \ldots, n$. Como cada una de las Y_i está distribuida exponencialmente con media θ, $m_{Y_i}(t) = (1 - \theta t)^{-1}$ y, por el Teorema 6.2,

$$\begin{aligned} m_U(t) &= m_{Y_1}(t) \times m_{Y_1}(t) \times \cdots \times m_{Y_n}(t) \\ &= (1-\theta t)^{-1} \times (1-\theta t)^{-1} \times \cdots \times (1-\theta t)^{-1} = (1-\theta t)^{-n}. \end{aligned}$$

Ésta es la función generadora de momento de una variable aleatoria con distribución gamma y $\alpha = n$ y $\beta = \theta$. El Teorema 6.1 implica que U en realidad tiene esta distribución gamma y por tanto que

$$f_U(u) = \begin{cases} \dfrac{1}{\Gamma(n)\theta^n}\left(u^{n-1}e^{-u/\theta}\right), & u > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

■

El método de las funciones generadoras de momento se puede usar para establecer algunos resultados útiles e interesantes acerca de las distribuciones de funciones de variables aleatorias normalmente distribuidas. Como estos resultados se usarán en los Capítulos 7~9, los presentamos en la forma de teoremas.

TEOREMA 6.3 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $V(Y_i) = \sigma_i^2$, para $i = 1, 2, \ldots, n$, y sean $a_1, a_2, \ldots, a_n$ constantes. Si

$$U = \sum_{i=1}^{n} a_i Y_i = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n,$$

entonces U es una variable aleatoria normalmente distribuida con

$$E(U) = \sum_{i=1}^{n} a_i \mu_i = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

y

$$V(U) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2.$$

Demostración Como Y_i está normalmente distribuida con media μ_i y varianza σ_i^2, Y_i tiene función generadora de momento dada por

$$m_{Y_i}(t) = \exp\left(\mu_i t + \frac{\sigma_i^2 t^2}{2}\right).$$

[Recuerde que exp(·) es una forma más cómoda de escribir $e^{(\cdot)}$ cuando el término en el exponente es largo o complejo.] Por tanto, $a_i Y_i$ tiene función generadora de momento dada por

$$m_{a_i Y_i}(t) = E(e^{t a_i Y_i}) = m_{Y_i}(a_i t) = \exp\left(\mu_i a_i t + \frac{a_i^2 \sigma_i^2 t^2}{2}\right).$$

Debido a que las variables aleatorias Y_i son independientes, las variables aleatorias $a_i Y_i$ son independientes, para $i = 1, 2, \ldots, n$, y el Teorema 6.2 implica que

$$\begin{aligned} m_U(t) &= m_{a_1 Y_1}(t) \times m_{a_2 Y_2}(t) \times \cdots \times m_{a_n Y_n}(t) \\ &= \exp\left(\mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2}\right) \times \cdots \times \exp\left(\mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2}\right) \\ &= \exp\left(t \sum_{i=1}^{n} a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2\right). \end{aligned}$$

Entonces, U tiene una distribución normal con media $\sum_{i=1}^{n} a_i \mu_i$ y varianza $\sum_{i=1}^{n} a_i^2 \sigma_i^2$.

TEOREMA 6.4 Sean $Y_1, Y_2, \ldots, Y_n$ definidas como en el Teorema 6.3 y definimos Z_i por

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}, \quad i = 1, 2, \ldots, n.$$

Entonces $\sum_{i=1}^{n} Z_i^2$ tiene una distribución χ^2 con n grados de libertad.

Demostración Como Y_i está normalmente distribuida con media μ_i y varianza σ_i^2, el resultado del Ejemplo 6.10 implica que Z_i está normalmente distribuida con media 0 y varianza 1. Del Ejemplo 6.11, entonces tenemos que Z_i^2 es una variable aleatoria con distribución χ^2 con 1 grado de libertad. En consecuencia,

$$m_{Z_i^2}(t) = (1-2t)^{-1/2},$$

y del Teorema 6.2, con $V = \sum_{i=1}^{n} Z_i^2$,

$$\begin{aligned} m_V(t) &= m_{Z_1^2}(t) \times m_{Z_2^2}(t) \times \cdots \times m_{Z_n^2}(t) \\ &= (1-2t)^{-1/2} \times (1-2t)^{-1/2} \times \cdots \times (1-2t)^{-1/2} = (1-2t)^{-n/2}. \end{aligned}$$

Como las funciones generadoras de momento son únicas, V tiene una distribución χ^2 con n grados de libertad.

El teorema 6.4 proporciona alguna aclaración de los *grados de libertad* asociados con una distribución χ^2. Si n es independiente, las variables aleatorias normales estándar se elevan al cuadrado y se suman y el resultado tiene una distribución χ^2 con n grados de libertad.

Resumen del método de las funciones generadoras de momento

Sea U una función de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$.

1. Encuentre la función generadora de momento para U, $m_U(t)$.
2. Compare $m_U(t)$ con otras funciones generadoras de momento bien conocidas. Si $m_U(t) = m_V(t)$ para todos los valores de t, el Teorema 6.1 implica que U y V tienen distribuciones idénticas.

Ejercicios

6.37 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes e idénticamente distribuidas tales que para $0 < p < 1$, $P(Y_i = 1) = p$ y $P(Y_i = 0) = q = 1 - p$. (Tales variables aleatorias reciben el nombre de variables aleatorias de *Bernoulli*.)

a Encuentre la función generadora de momento para la variable aleatoria Y_1 de Bernoulli.
b Encuentre la función generadora de momento para $W = Y_1 + Y_2 + \cdots + Y_n$.
c ¿Cuál es la distribución de W?

6.38 Sean Y_1 y Y_2 variables aleatorias independientes con funciones generadoras de momento $m_{Y_1}(t)$ y $m_{Y_2}(t)$, respectivamente. Si a_1 y a_2 son constantes, y $U = a_1Y_1 + a_2Y_2$ demuestre que la función generadora de momento para U es $m_U(t) = m_{Y_1}(a_1t) \times m_{Y_2}(a_2t)$.

6.39 En los Ejercicios 6.11 y 6.25 consideramos dos componentes electrónicos que operan independientemente, cada uno con una vida útil gobernada por la distribución exponencial con media 1. Use el método de las funciones generadoras de momento para obtener la función de densidad para el promedio de vida útil de los dos componentes.

6.40 Suponga que Y_1 y Y_2 son variables aleatorias normales estándar e independientes. Encuentre la función de densidad de $U = Y_1^2 + Y_2^2$.

6.41 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias normales e independientes, cada una con media μ y varianza σ^2. Denote con $a_1, a_2, \ldots, a_n$ constantes conocidas. Encuentre la función de densidad de la combinación lineal $U = \sum_{i=1}^{n} a_i Y_i$.

6.42 Un tipo de elevador tiene una capacidad máxima de peso Y_1, que está normalmente distribuida con media de 5000 libras y desviación estándar de 300 libras. Para un cierto edificio equipado con este tipo de elevador, la carga del elevador, Y_2, es una variable aleatoria normalmente distribuida con media de 4000 libras y desviación estándar de 400 libras. Para cualquier tiempo determinado en el que el elevador está en uso, calcule la probabilidad de que sea sobrecargado, suponiendo que Y_1 y Y_2 son independientes.

6.43 Consulte el Ejercicio 6.41. Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias normales e independientes, cada una con media μ y varianza σ^2.

a Encuentre la función de densidad de $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^{n} Y_i$.

b Si $\sigma^2 = 16$ y $n = 25$, ¿cuál es la probabilidad de que la media muestral, $\overline{Y}$, tome un valor que esté a no más de una unidad de la media poblacional, μ? Esto es, encuentre $P(|\overline{Y} - \mu| \le 1)$.

c Si $\sigma^2 = 16$, encuentre $P(|\overline{Y} - \mu| \le 1)$ si $n = 36$, $n = 64$ y $n = 81$. Interprete los resultados de sus cálculos.

***6.44** El peso (en libras) de sandías de "tamaño mediano" está normalmente distribuido con media de 15 y varianza de 4. Un recipiente de empaque para varias sandías tiene una capacidad nominal de 140 libras. ¿Cuál es el número máximo de sandías que deben ponerse en un solo recipiente de empaque si el límite de peso nominal debe excederse sólo el 5% del tiempo? Justifique su respuesta.

6.45 El gerente de una obra de construcción necesita hacer una cotización de precios con todo cuidado antes de presentar un presupuesto. También necesita tomar en cuenta la incertidumbre (variabilidad) en las cantidades de productos que podría necesitar. Para simplificar al máximo la situación real, suponga que un gerente de proyectos representa la cantidad de arena, en yardas, necesaria para un proyecto de construcción como si fuera una variable aleatoria Y_1, que está normalmente distribuida con media de 10 yardas y desviación estándar de .5 yarda. La cantidad de mezcla de cemento necesaria, en cientos de libras, es una variable aleatoria Y_2 que está normalmente distribuida con media de 4 y desviación estándar .2. La arena cuesta \$7 por yarda, y la mezcla de cemento cuesta \$3 por cien libras. Sumando \$100 por otros costos, el gerente calcula que su costo total es

$$U = 100 + 7Y_1 + 3Y_2.$$

Si Y_1 y Y_2 son independientes, ¿qué presupuesto debe presentar el gerente para asegurar que los costos reales rebasarán la cantidad cotizada con una probabilidad de sólo .01? ¿En este caso es razonable la suposición de independencia?

6.46 Suponga que Y tiene una distribución gamma con $\alpha = n/2$ para algún entero positivo n y β igual a algún valor especificado. Use el método de las funciones generadoras de momento para demostrar que $W = 2Y/\beta$ tiene una distribución χ^2 con n grados de libertad.

6.47 Una variable aleatoria Y tiene una distribución gamma con $\alpha = 3.5$ y $\beta = 4.2$. Use el resultado del Ejercicio 6.46 y los puntos porcentuales para las distribuciones χ^2 dadas en la Tabla 6, Apéndice 3, para hallar $P(Y > 33.627)$.

6.48 En un programa de prueba de misiles, una variable aleatoria de interés es la distancia entre el punto de impacto del misil y el centro del blanco al que fue dirigido el misil. Si consideramos el centro del blanco como el origen de un sistema de coordenadas, podemos denotar con Y_1 la distancia norte-sur entre el punto de impacto y el centro del blanco y denotar con Y_2 la correspondiente distancia este-oeste.

(Suponga que norte y este definen direcciones positivas.) La distancia entre el punto de impacto y el centro del blanco es entonces $U = \sqrt{Y_1^2 + Y_2^2}$. Si Y_1 y Y_2 son variables aleatorias normales estándar e independientes, encuentre la función de densidad de probabilidad para U.

6.49 Sea Y_1 una variable aleatoria binomial con n_1 intentos y probabilidad de éxito dada por p. Sea Y_2 otra variable aleatoria binomial con n_2 intentos y probabilidad de éxito también dada por p. Si Y_1 y Y_2 son independientes, encuentre la función de probabilidad de $Y_1 + Y_2$.

6.50 Sea Y una variable aleatoria binomial con n intentos y probabilidad de éxito dada por p. Demuestre que $n - Y$ es una variable aleatoria binomial con n intentos y probabilidad de éxito dada por $1 - p$.

6.51 Sea Y_1 una variable aleatoria binomial con n_1 intentos y $p_1 = .2$ y sea Y_2 una variable aleatoria binomial independiente con n_2 intentos y $p_2 = .8$. Encuentre la función de probabilidad de $Y_1 + n_2 - Y_2$.

6.52 Sean Y_1 y Y_2 variables aleatorias de Poisson independientes con medias λ_1 y λ_2, respectivamente. Encuentre

a la función de probabilidad de $Y_1 + Y_2$,
b la función de probabilidad condicional de Y_1, dado que $Y_1 + Y_2 = m$.

6.53 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias binomiales independientes con n_i intentos y probabilidad de éxito dada por p_i, $= 1, 2, \ldots, n$.

a Si todas las n_i son iguales y todas las p son iguales, encuentre la distribución de $\sum_{i=1}^{n} Y_i$.
b Si todas las n_i son diferentes y todas las p son iguales, encuentre la distribución de $\sum_{i=1}^{n} Y_i$.
c Si todas las n_i son diferentes y todas las p son iguales, encuentre la distribución condicional Y_1 dado que $\sum_{i=1}^{n} Y_i = m$.
d Si todas las n_i son diferentes y todas las p son iguales, encuentre la distribución condicional $Y_1 + Y_2$ dado que $\sum_{i=1}^{n} Y_i = m$.
e Si todas las p son diferentes, ¿el método de las funciones generadoras de momento funciona bien para encontrar la distribución de $\sum_{i=1}^{n} Y_i$? ¿Por qué?

6.54 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes de Poisson con medias $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectivamente. Encuentre

a la función de probabilidad de $\sum_{i=1}^{n} Y_i$.
b la función de probabilidad condicional de Y_1, dado que $\sum_{i=1}^{n} Y_i = m$.
c la función de probabilidad condicional de $Y_1 + Y_2$, dado que $\sum_{i=1}^{n} Y_i = m$.

6.55 Llegan clientes a la caja de una tienda departamental de acuerdo con una distribución de Poisson, con media de 7 por hora. En un periodo determinado de dos horas, ¿cuál es la probabilidad de que 20 o más clientes lleguen a la caja?

6.56 El tiempo necesario para afinar un automóvil está exponencialmente distribuido con una media de .5 hora. Si dos autos están en espera de una afinación y los tiempos de servicio son independientes, ¿cuál es la probabilidad de que el tiempo total para afinar los dos automóviles sea mayor que 1.5 horas? [*Sugerencia:* recuerde el resultado del Ejemplo 6.12.]

6.57 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes tales que cada Y_i tiene una distribución gamma con parámetros α_i y β. Esto es, las distribuciones de las Y podrían tener diferentes α, pero todas tienen el mismo valor para β. Demuestre que $U = Y_1 + Y_2 + \cdots + Y_n$ tiene una distribución gamma con parámetros $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ y β.

6.58 Vimos en el Ejercicio 5.159 que la variable aleatoria binomial negativa Y se puede escribir como $Y = \sum_{i=1}^{r} W_i$, donde $W_1, W_2, \ldots, W_r$ son variables aleatorias geométricas independientes con parámetro p.

a Use este dato para obtener la función generadora de momento para Y.

b Use la función generadora de momento para demostrar que $E(Y) = r/p$ y $V(Y) = r(1 - p)/p^2$.

c Encuentre la función de probabilidad condicional para W_1, dado que $Y = W_1 + W_2 + \cdots + W_r = m$.

6.59 Demuestre que si Y_1 tiene una distribución χ^2 con ν_1 grados de libertad y Y_2 tiene una distribución χ^2 con ν_2 grados de libertad, entonces $U = Y_1 + Y_2$ tiene una distribución χ^2 con $\nu_1 + \nu_2$ grados de libertad, siempre que Y_1 y Y_2 sean independientes.

6.60 Suponga que $W = Y_1 + Y_2$ donde Y_1 y Y_2 son independientes. Si W tiene una distribución χ^2 con ν grados de libertad y W_1 tiene una distribución χ^2 con $\nu_1 < \nu$ grados de libertad, demuestre que Y_2 tiene una distribución χ^2 con $\nu - \nu_1$ grados de libertad.

6.61 Consulte el Ejercicio 6.52. Suponga que $W = Y_1 + Y_2$ donde Y_1 y Y_2 son independientes. Si W tiene una distribución de Poisson con media λ y W_1 tiene una distribución de Poisson con media $\lambda_1 < \lambda$, demuestre que Y_2 tiene una distribución de Poisson con media $\lambda - \lambda_1$.

***6.62** Sean Y_1 y Y_2 variables aleatorias normales independientes, cada una con media 0 y varianza σ^2. Defina $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$. Demuestre que U_1 y U_2 son variables aleatorias normales independientes, cada una con media 0 y varianza $2\sigma^2$. [*Sugerencia:* si (U_1, U_2) tiene una función generadora de momento conjunta $m(t_1, t_2)$, entonces U_1 y U_2 son independientes si y sólo si $m(t_1, t_2) = m_{U_1}(t_1)m_{U_2}(t_2)$.]

6.6 Transformaciones multivariantes usando jacobianos (opcional)

Si Y es una variable aleatoria con función de densidad $f_Y(y)$, el método de las transformaciones (Sección 6.4) se puede usar para determinar la función de densidad para $U = h(Y)$, siempre que $h(y)$ sea creciente o decreciente para toda y tal que $f_Y(y) > 0$. Si $h(y)$ es creciente o decreciente para toda y en el soporte de $f_Y(y)$, la función $h(\cdot)$ es uno a uno y existe una función inversa $h^{-1}(\cdot)$ tal que $u = h^{-1}(y)$. Además, la función de densidad para U está dada por

$$f_U(u) = f_Y(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right|.$$

Suponga que Y_1 y Y_2 son variables aleatorias continuas conjuntas y que $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$. ¿Cómo podemos determinar la *función de densidad conjunta* de U_1 y U_2?

En el resto de esta sección expresaremos la densidad conjunta de Y_1 y Y_2 como $f_{Y_1,Y_2}(y_1, y_2)$. Extendiendo las ideas de la Sección 6.4, el soporte de la densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ es el conjunto de todos los valores de (y_1, y_2) tales que $f_{Y_1,Y_2}(y_1, y_2) > 0$.

Método de las transformaciones bivariadas

Suponga que Y_1 y Y_2 son variables aleatorias continuas con función de densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ y que para toda (y_1, y_2) tal que $f_{Y_1,Y_2}(y_1, y_2) > 0$.

$$u_1 = h_1(y_1, y_2) \quad \text{y} \quad u_2 = h_2(y_1, y_2)$$

es una transformación uno a uno de (y_1, y_2) y (u_1, u_2) con inversa

$$y_1 = h_1^{-1}(u_1, u_2) \quad \text{y} \quad y_2 = h_2^{-1}(u_1, u_2).$$

Si $h_1^{-1}(u_1, u_2)$ y $h_2^{-1}(u_1, u_2)$ tienen derivadas parciales continuas con respecto a u_1 y u_2 y jacobiano

$$J = \det \begin{bmatrix} \dfrac{\partial h_1^{-1}}{\partial u_1} & \dfrac{\partial h_1^{-1}}{\partial u_2} \\ \dfrac{\partial h_2^{-1}}{\partial u_1} & \dfrac{\partial h_2^{-1}}{\partial u_2} \end{bmatrix} = \frac{\partial h_1^{-1}}{\partial u_1}\frac{\partial h_2^{-1}}{\partial u_2} - \frac{\partial h_2^{-1}}{\partial u_1}\frac{\partial h_1^{-1}}{\partial u_2} \neq 0,$$

entonces la densidad conjunta de U_1 y U_2 es

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}\left(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)\right)|J|,$$

donde $|J|$ es el valor absoluto del jacobiano.

No demostraremos este resultado, pero se deduce de resultados de cálculo empleados para cambio de variables en la integración múltiple. (Recuerde que a veces es más fácil calcular las integrales dobles si usamos coordenadas polares en lugar de coordenadas euclidianas; véase el Ejercicio 4.194.) El valor absoluto del jacobiano, $|J|$, en la transformación multivariante es análogo a la cantidad $|dh^{-1}(u)/du|$ que se usa cuando se hace la transformación de una variable $U = h(Y)$.

Advertencia. *Es necesario asegurarse que la transformación bivariante* $u_1 = h_1(y_1, y_2)$, $u_2 = h_2(y_1, y_2)$ *es una transformación biunívoca para toda* (y_1, y_2) *tal que* $f_{Y_1,Y_2}(y_1, y_2) > 0$. Es frecuente omitir este paso. Si la transformación bivariante no es biunívoca y este método se aplica a ciegas, la función resultante de "densidad" no tendrá las propiedades necesarias de una función de densidad válida. Ilustramos el uso de este método en los ejemplos siguientes.

EJEMPLO **6.13** Sean Y_1 y Y_2 variables aleatorias normales estándar independientes. Si $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$, entonces U_1 y U_2 son combinaciones lineales de variables aleatorias normalmente distribuidas e independientes, y el Teorema 6.3 implica que U_1 está normalmente distribuida con media $0 + 0 = 0$ y varianza $1 + 1 = 2$. Del mismo modo, U_2 tiene una distribución normal con media 0 y varianza 2. ¿Cuál es la *densidad conjunta* de U_1 y U_2?

Solución Las funciones de densidad Y_1 y Y_2 son

$$f_1(y_1) = \frac{e^{-(1/2)y_1^2}}{\sqrt{2\pi}}, \qquad -\infty < y_1 < \infty$$

$$f_2(y_2) = \frac{e^{-(1/2)y_2^2}}{\sqrt{2\pi}}, \qquad -\infty < y_2 < \infty,$$

y la independencia de Y_1 y Y_2 implica que la densidad conjunta de ambas es

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{e^{-(1/2)y_1^2-(1/2)y_2^2}}{2\pi}, \qquad -\infty < y_1 < \infty, \ -\infty < y_2 < \infty.$$

En este caso, $f_{Y_1,Y_2}(y_1, y_2) > 0$ para toda $-\infty < y_1 < \infty$ y $-\infty < y_2 < \infty$, y estamos interesados en la transformación

$$u_1 = y_1 + y_2 = h_1(y_1, y_2) \quad \text{y} \quad u_2 = y_1 - y_2 = h_2(y_1, y_2)$$

con transformación inversa

$$y_1 = (u_1 + u_2)/2 = h_1^{-1}(u_1, u_2) \text{ y } y_2 = (u_1 - u_2)/2 = h_2^{-1}(u_1, u_2).$$

Como $\partial h_1^{-1}/\partial u_1 = 1/2$, $\partial h_1^{-1}/\partial u_2 = 1/2$, $\partial h_2^{-1}/\partial u_1 = 1/2$ y $\partial h_2^{-1}/\partial u_2 = -1/2$, el jacobiano de esta transformación es

$$J = \det\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} = (1/2)(-1/2) - (1/2)(1/2) = -1/2$$

y la densidad conjunta de U_1 y U_2 es [con $\exp(\cdot) = e^{(\cdot)}$]

$$f_{U_1,U_2}(u_1, u_2) = \frac{\exp\left[-\frac{1}{2}\left(\frac{u_1+u_2}{2}\right)^2 - \frac{1}{2}\left(\frac{u_1-u_2}{2}\right)^2\right]}{2\pi}\left|-\frac{1}{2}\right|, \quad \begin{array}{l} -\infty < (u_1 + u_2)/2 < \infty, \\ -\infty < (u_1 - u_2)/2 < \infty. \end{array}$$

Un poco de álgebra nos dará como resultado

$$-\frac{1}{2}\left(\frac{u_1 + u_2}{2}\right)^2 - \frac{1}{2}\left(\frac{u_1 - u_2}{2}\right)^2 = -\frac{1}{4}u_1^2 - \frac{1}{4}u_2^2$$

y

$$\begin{aligned} &\{(u_1, u_2) : -\infty < (u_1 + u_2)/2 < \infty, -\infty < (u_1 - u_2)/2 < \infty\} \\ &= \{(u_1, u_2) : -\infty < u_1 < \infty, -\infty < u_2 < \infty\}. \end{aligned}$$

Finalmente, como $4\pi = \sqrt{2}\sqrt{2\pi}\sqrt{2}\sqrt{2\pi}$,

$$f_{U_1,U_2}(u_1, u_2) = \frac{e^{-u_1^2/4}}{\sqrt{2}\sqrt{2\pi}}\frac{e^{-u_2^2/4}}{\sqrt{2}\sqrt{2\pi}}, \qquad -\infty < u_1 < \infty,\ -\infty < u_2 < \infty.$$

Observe que U_1 y U_2 son *independientes* y normalmente distribuidas, ambas con media 0 y varianza 2. ¡La información extra dada por la distribución conjunta de U_1 y U_2 es que las dos variables son independientes! ■

El método de las transformaciones multivariantes también es útil si estamos interesados en una sola función de Y_1 y Y_2, por ejemplo, $U_1 = h(Y_1, Y_2)$. Como tenemos sólo una función de Y_1 y Y_2, podemos usar el método de transformaciones bivariantes para hallar la distribución *conjunta* de U_1 y otra función $U_2 = h_2(Y_1, Y_2)$ y en seguida determinar la densidad marginal deseada de U_1 al integrar la densidad conjunta. Debido a que estamos interesados realmente sólo en la distribución de U_1, generalmente escogeríamos la otra función $U_2 = h_2(Y_1, Y_2)$ para que la transformación bivariante sea fácil de invertir y sea sencillo trabajar con el jacobiano. Ilustramos la aplicación de esta técnica en el siguiente ejemplo.

EJEMPLO **6.14** Sean Y_1 y Y_2 variables aleatorias exponenciales independientes, ambas con media $\beta > 0$. Encuentre la función de densidad de

$$U = \frac{Y_1}{Y_1 + Y_2}.$$

Solución Las funciones de densidad para Y_1 y Y_2 son, de nuevo usando $\exp(\cdot) = e^{(\cdot)}$,

$$f_1(y_1) = \begin{cases} \dfrac{1}{\beta}\exp(-y_1/\beta), & 0 < y_1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y

$$f_2(y_2) \begin{cases} \dfrac{1}{\beta}\exp(-y_2/\beta), & 0 < y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Su densidad conjunta es

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{1}{\beta^2}\exp[-(y_1 + y_2)/\beta], & 0 < y_1,\ 0 < y_2, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

porque Y_1 y Y_2 son independientes.

En este caso, $f_{Y_1,Y_2}(y_1, y_2) > 0$ para toda (y_1, y_2) tal que $0 < y_1$, $0 < y_2$, y estamos interesados en la función $U_1 = Y_1/(Y_1 + Y_2)$. Si consideramos la función $u_1 = y_1/(y_1 + y_2)$, obviamente hay muchos valores para (y_1, y_2) que darán el mismo valor para u_1. Definamos

$$u_1 = \frac{y_1}{y_1 + y_2} = h_1(y_1, y_2) \quad \text{y} \quad u_2 = y_1 + y_2 = h_2(y_1, y_2).$$

Elegir esta opción de u_2 da como resultado una transformación inversa adecuada:

$$y_1 = u_1u_2 = h_1^{-1}(u_1, u_2) \quad \text{y} \quad y_2 = u_2(1 - u_1) = h_2^{-1}(u_1, u_2).$$

El jacobiano de esta transformación es

$$J = \det\begin{bmatrix} u_2 & u_1 \\ -u_2 & 1 - u_1 \end{bmatrix} = u_2(1 - u_1) - (-u_2)(u_1) = u_2,$$

y la densidad conjunta de U_1 y U_2 es

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} \dfrac{1}{\beta^2}\exp\{-[u_1u_2 + u_2(1 - u_1)]/\beta\}\left|u_2\right|, & 0 < u_1u_2,\ 0 < u_2(1 - u_1), \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En este caso $f_{U_1,U_2}(u_1, u_2) > 0$ si u_1 y u_2 son tales que $0 < u_1u_2$, $0 < u_2(1 - u_1)$. Observe que si $0 < u_1u_2$, entonces

$$0 < u_2(1 - u_1) = u_2 - u_1u_2 \quad \Leftrightarrow \quad 0 < u_1u_2 < u_2 \quad \Leftrightarrow \quad 0 < u_1 < 1.$$

Si $0 < u_1 < 1$, entonces $0 < u_2(1 - u_1)$ implica que $0 < u_2$. Por tanto, la región de soporte para la densidad conjunta de U_1 y U_2 es $\{(u_1, u_2): 0 < u_1 < 1, 0 < u_2\}$, y la densidad conjunta de U_1 y U_2 está dada por

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} \dfrac{1}{\beta^2}u_2e^{-u_2/\beta}, & 0 < u_1 < 1,\ 0 < u_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Usando el Teorema 5.5 se ve fácilmente que U_1 y U_2 son independientes. Las densidades marginales de U_1 y U_2 se pueden obtener al integrar la densidad conjunta que se obtuvo antes.

En el Ejercicio 6.63 usted demostrará que U_1 está uniformemente distribuida en (0, 1) y que U_2 tiene una densidad gamma con parámetros $\alpha = 2$ y β. ■

La técnica descrita en esta sección se puede ver como una versión de un paso del proceso de dos pasos ilustrado en el Ejemplo 6.9.

En el Ejemplo 6.14 fue más difícil hallar la región de soporte (donde la densidad conjunta es positiva) que la ecuación de la función de densidad conjunta. Como veremos en el ejemplo y ejercicios siguientes, con frecuencia éste es el caso.

EJEMPLO 6.15 En el Ejemplo 6.9, consideramos las variables aleatorias Y_1 y Y_2 con función de densidad conjunta

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} 2(1 - y_1), & 0 \le y_1 \le 1,\ 0 \le y_2 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y nos interesaba $U = Y_1Y_2$. Encuentre la función de densidad de probabilidad para U usando el método de transformación bivariante.

Solución En este caso $f_{Y_1,Y_2}(y_1, y_2) > 0$ para toda (y_1, y_2), tales que $0 \le y_1 \le 1$, $0 \le y_2 \le 1$, y estamos interesados en la función $U_2 = Y_1Y_2$. Si consideramos la función $u_2 = y_1y_2$, por sí sola no es una función biunívoca de las variables (y_1, y_2). Considere

$$u_1 = y_1 = h_1(y_1, y_2) \quad \text{y} \quad u_2 = y_1y_2 = h_2(y_1, y_2).$$

Para esta elección de u_1, y $0 \le y_1 \le 1$, $0 \le y_2 \le 1$, la transformación de (y_1, y_2) en (u_1, u_2) es biunívoca y

$$y_1 = u_1 = h_1^{-1}(u_1, u_2) \quad \text{y} \quad y_2 = u_2/u_1 = h_2^{-1}(u_1, u_2).$$

El jacobiano es

$$J = \det\begin{bmatrix} 1 & 0 \\ -u_2/u_1^2 & 1/u_1 \end{bmatrix} = 1(1/u_1) - (-u_2/u_1^2)(0) = 1/u_1.$$

La variable original de interés es $U_2 = Y_1Y_2$, y la densidad conjunta de U_1 y U_2 es

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} 2(1 - u_1)\left|\dfrac{1}{u_1}\right|, & 0 \le u_1 \le 1,\ 0 \le u_2/u_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Debido a que

$$\{(u_1, u_2)\colon 0 \le u_1 \le 1,\ 0 \le u_2/u_1 \le 1\} = \{(u_1, u_2)\colon 0 \le u_2 \le u_1 \le 1\},$$

la densidad conjunta de U_1 y U_2 es

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} 2(1 - u_1)\dfrac{1}{u_1}, & 0 \le u_2 \le u_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Esta densidad conjunta es exactamente igual que la densidad conjunta obtenida en el Ejemplo 6.9 si identificamos las variables Y_1 y U empleadas en el Ejemplo 6.9 con las variables U_1 y

U_2, respectivamente, usadas aquí. Con esta identificación, la densidad marginal de U_2 es precisamente la densidad de U obtenida en el Ejemplo 6.9, es decir,

$$f_2(u_2) = \begin{cases} 2(u_2 - \ln u_2 - 1), & 0 \leq u_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$ ■

Si $Y_1, Y_2, \ldots, Y_k$ son variables aleatorias continuas conjuntas y

$$U_1 = h_1(Y_1, Y_2, \ldots, Y_k), \; U_2 = h_2(Y_1, Y_2, \ldots, Y_k), \ldots, \; U_k = h_k(Y_1, Y_2, \ldots, Y_k),$$

donde la transformación

$$u_1 = h_1(y_1, y_2, \ldots, y_k), u_2 = h_2(y_1, y_2, \ldots, y_k), \ldots, u_k = h_k(y_1, y_2, \ldots, y_k)$$

es una transformación biunívoca de $(y_1, y_2, \ldots, y_k)$ en $(u_1, u_2, \ldots, u_k)$ con inversa

$$y_1 = h_1^{-1}(u_1, u_2, \ldots, u_k), \; y_2 = h_2^{-1}(u_1, u_2, \ldots, u_k), \ldots,$$
$$y_k = h_k^{-1}(u_1, u_2, \ldots, u_k),$$

y $h_1^{-1}(u_1, u_2, \ldots, u_k), \; h_2^{-1}(u_1, u_2, \ldots, u_k), \ldots, \; h_k^{-1}(u_1, u_2, \ldots, u_k)$ tienen derivadas parciales continuas con respecto a $u_1, u_2, \ldots, u_k$ y *jacobiano*

$$J = \det \begin{bmatrix} \frac{\partial h_1^{-1}}{\partial u_1} & \frac{\partial h_1^{-1}}{\partial u_2} & \cdots & \frac{\partial h_1^{-1}}{\partial u_k} \\ \frac{\partial h_2^{-1}}{\partial u_1} & \frac{\partial h_2^{-1}}{\partial u_2} & \cdots & \frac{\partial h_2^{-1}}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_k^{-1}}{\partial u_1} & \frac{\partial h_k^{-1}}{\partial u_2} & \cdots & \frac{\partial h_k^{-1}}{\partial u_k} \end{bmatrix} \neq 0,$$

entonces se puede usar un resultado similar al presentado en esta sección para determinar la densidad conjunta de $U_1, U_2, \ldots, U_k$. Esto requiere calcular el determinante de una matriz $k \times k$, este conocimiento no se requiere en el resto de este texto. Para más detalles, vea "Bibliografía y lecturas adicionales" al final del capítulo.

Ejercicios

***6.63** En el Ejemplo 6.14, Y_1 y Y_2 eran variables aleatorias independientes distribuidas exponencialmente, ambas con media β. Definimos $U_1 = Y_1/(Y_1 + Y_2)$ y $U_2 = Y_1 + Y_2$ y determinamos la densidad conjunta de (U_1, U_2) como

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} \frac{1}{\beta^2} u_2 e^{-u_2/\beta}, & 0 < u_1 < 1, \; 0 < u_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que U_1 está uniformemente distribuida en el intervalo (0, 1).

b Demuestre que U_2 tiene una densidad gamma con parámetros $\alpha = 2$ y β.

c Establezca que U_1 y U_2 son independientes.

***6.64** Consulte el Ejercicio 6.63 y el Ejemplo 6.14. Suponga que Y_1 tiene una distribución gamma con parámetros α_1 y β, que Y_1 tiene distribución gamma con parámetros α_2 y β, y que Y_1 y Y_2 son independientes. Sea $U_1 = Y_1/(Y_1 + Y_2)$ y $U_2 = Y_1 + Y_2$.

a Deduzca la función de densidad conjunta para U_1 y U_2.

b Demuestre que la distribución marginal de U_1 es una distribución beta con parámetros α_1 y α_2.

c Demuestre que la distribución marginal de U_2 es una distribución gamma con parámetros $\alpha = \alpha_1 + \alpha_2$ y β.

d Establezca que U_1 y U_2 son independientes.

6.65 Sean Z_1 y Z_2 variables aleatorias normales estándar independientes y $U_1 = Z_1$ y $U_2 = Z_1 + Z_2$.

a Deduzca la densidad conjunta de U_1 y U_2.

b Use el Teorema 5.12 para obtener $E(U_1)$, $E(U_2)$, $V(U_1)$, $V(U_2)$ y $\text{Cov}(U_1, U_2)$.

c ¿U_1 y U_2 son independientes? ¿Por qué?

d Consulte la Sección 5.10. Demuestre que U_1 y U_2 tienen una distribución normal bivariada. Identifique todos los parámetros de la distribución normal bivariada apropiada.

***6.66** Sean (Y_1, Y_2) que tienen función de densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ y sean $U_1 = Y_1 + Y_2$ y $U_2 = Y_2$.

a Demuestre que la densidad conjunta de (U_1, U_2) es

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}(u_1 - u_2, u_2).$$

b Demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(u_1 - u_2, u_2)\, du_2.$$

c Si Y_1 y Y_2 son independientes, demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1}(u_1 - u_2) f_{Y_2}(u_2)\, du_2.$$

Esto es, que la densidad de $Y_1 + Y_2$ es la *convolución* de las densidades $f_{Y_1}(\cdot)$ y $f_{Y_2}(\cdot)$.

***6.67** Sean (Y_1, Y_2) que tienen función de densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ y sean $U_1 = Y_1/Y_2$ y $U_2 = Y_2$.

a Demuestre que la densidad conjunta de (U_1, U_2) es

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}(u_1 u_2, u_2)|u_2|.$$

b Demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(u_1 u_2, u_2)|u_2|\, du_2.$$

c Si Y_1 y Y_2 son independientes, demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1}(u_1 u_2) f_{Y_2}(u_2)|u_2|\, du_2.$$

***6.68** Sean Y_1 y Y_2 que tienen la función de densidad conjunta

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} 8y_1y_2, & 0 \le y_1 < y_2 \le 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y $U_1 = Y_1/Y_2$ y $U_2 = Y_2$.

a Deduzca la función de densidad conjunta para (U_1, U_2).

b Demuestre que U_1 y U_2 son independientes.

***6.69** Las variables aleatorias Y_1 y Y_2 son independientes, ambas con densidad

$$f(y) = \begin{cases} \dfrac{1}{y^2}, & 1 < y, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Sean $U_1 = \dfrac{Y_1}{Y_1 + Y_2}$ y $U_2 = Y_1 + Y_2$.

a ¿Cuál es la densidad conjunta de Y_1 y Y_2?

b Demuestre que la densidad conjunta de U_1 y U_2 está dada por

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} \dfrac{1}{u_1^2(1-u_1)^2u_2^3}, & 1/u_1 < u_2,\ 0 < u_1 < 1/2 \text{ y} \\ & 1/(1-u_1) < u_2,\ 1/2 \le u_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

c Dibuje la región donde $f_{U_1, U_2}(u_1, u_2) > 0$.

d Demuestre que la densidad marginal de U_1 es

$$f_{U_1}(u_1) = \begin{cases} \dfrac{1}{2(1-u_1)^2}, & 0 \le u_1 < 1/2, \\ \dfrac{1}{2u_1^2}, & 1/2 \le u_1 \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

e ¿U_1 y U_2 son independientes? ¿Por qué sí o por qué no?

***6.70** Suponga que Y_1 y Y_2 son independientes y que ambas están uniformemente distribuidas en el intervalo $(0, 1)$, y sean $U_1 = Y_1 + Y_2$ y $U_2 = Y_1 - Y_2$.

a Demuestre que la densidad conjunta de U_1 y U_2 está dada por

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} 1/2, & -u_1 < u_2 < u_1,\ 0 < u_1 < 1 \text{ y} \\ & u_1 - 2 < u_2 < 2 - u_1,\ 1 \le u_1 < 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b Dibuje la región donde $f_{U_1, U_2}(u_1, u_2) > 0$.

c Demuestre que la densidad marginal de U_1 es

$$f_{U_1}(u_1) = \begin{cases} u_1, & 0 < u_1 < 1, \\ 2 - u_1, & 1 \le u_1 < 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

d Demuestre que la densidad marginal de U_2 es

$$f_{U_2}(u_2) = \begin{cases} 1 + u_2, & -1 < u_2 < 0, \\ 1 - u_2, & 0 \le u_1 < 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

e ¿U_1 y U_2 son independientes? ¿Por qué sí o por qué no?

***6.71** Suponga que Y_1 y Y_2 son variables aleatorias e independientes distribuidas exponencialmente, ambas con media β y defina $U_1 = Y_1 + Y_2$ y $U_2 = Y_1/Y_2$.

a Demuestre que la densidad conjunta de (U_1, U_2) es

$$f_{U_1,U_2}(u_1, u_2) = \begin{cases} \dfrac{1}{\beta^2} u_1 e^{-u_1/\beta} \dfrac{1}{(1+u_2)^2}, & 0 < u_1,\ 0 < u_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b ¿U_1 y U_2 son independientes? ¿Por qué?

6.7 Estadísticos de orden

Numerosas funciones de variables aleatorias de interés en la práctica dependen de las magnitudes relativas de las variables observadas. Por ejemplo, podemos estar interesados en la máxima velocidad que se alcanza en una carrera de automóviles o en el ratón más pesado de entre los alimentados con cierta dieta. Esto quiere decir que con frecuencia ordenamos variables aleatorias observadas de acuerdo con sus magnitudes. Las variables ordenadas resultantes se denominan *estadísticos de orden.*

Formalmente, denotemos con $Y_1, Y_2, \ldots, Y_n$ a variables aleatorias continuas e independientes, con función de distribución $F(y)$ y función de densidad $f(y)$. Denotamos las variables aleatorias ordenadas Y_i por $Y_{(1)}$ y $Y_{(2)}, \ldots, Y_{(n)}$, donde $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$. (Debido a que las variables aleatorias son continuas, los signos de igualdad pueden ignorarse.) Usando esta notación,

$$Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$$

es la mínima de las variables aleatorias Y_i, y

$$Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$$

es la máxima de las variables aleatorias Y_i.

Las funciones de densidad de probabilidad para $Y_{(1)}$ y $Y_{(n)}$ se pueden determinar usando el método de las funciones de distribución. Primero vamos a deducir la función de densidad de $Y_{(n)}$. Como $Y_{(n)}$ es la máxima de $Y_1, Y_2, \ldots, Y_n$, el evento $(Y_{(n)} \leq y)$ ocurrirá si y sólo si los eventos $(Y_i \leq y)$ ocurren para toda $i = 1, 2, \ldots, n$. Esto es,

$$P(Y_{(n)} \leq y) = P(Y_1 \leq y, Y_2 \leq y, \ldots, Y_n \leq y).$$

Debido a que las Y_i son independientes y $P(Y_i \leq y) = F(y)$ para $i = 1, 2, \ldots, n$, se deduce que la función de distribución de $Y_{(n)}$ está dada por

$$F_{Y_{(n)}}(y) = P(Y_{(n)} \leq y) = P(Y_1 \leq y)P(Y_2 \leq y) \cdots P(Y_n \leq y) = [F(y)]^n.$$

Si con $g_{(n)}(y)$ denotamos la función de densidad de $Y_{(n)}$, vemos que, al evaluar las derivadas de ambos lados,

$$g_{(n)}(y) = n[F(y)]^{n-1} f(y).$$

La función de densidad para $Y_{(1)}$ se puede hallar de un modo similar. La función de distribución de $Y_{(1)}$ es

$$F_{Y_{(1)}}(y) = P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} > y).$$

Como $Y_{(1)}$ es la mínima de $Y_1, Y_2, \ldots, Y_n$, se deduce que el evento $(Y_{(1)} > y)$ ocurre si y sólo si los eventos $(Y_i > y)$ ocurren para $i = 1, 2, \ldots, n$. Debido a que las Y_i son independientes y

$P(Y_i > y = 1 - F(y)$ para $i = 1, 2, \ldots, n$, vemos que

$$\begin{aligned} F_{Y_{(1)}}(y) &= P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} > y) \\ &= 1 - P(Y_1 > y, Y_2 > y, \ldots, Y_n > y) \\ &= 1 - [P(Y_1 > y)P(Y_2 > y) \cdots P(Y_n > y)] \\ &= 1 - [1 - F(y)]^n. \end{aligned}$$

Por consiguiente, si $g_{(1)}(y)$ denota la función de densidad de $Y_{(1)}$, al derivar en ambos lados de la última expresión obtenemos

$$g_{(1)}(y) = n[1 - F(y)]^{n-1} f(y).$$

Consideremos ahora el caso $n = 2$ y determinemos la función de densidad conjunta para $Y_{(1)}$ y $Y_{(2)}$. El evento $(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2)$ significa que $(Y_1 \leq y_1, Y_2 \leq y_2)$ o $(Y_2 \leq y_1, Y_1 \leq y_2)$. [Observe que $Y_{(1)}$ podría ser Y_1 o Y_2, cualquiera que sea más pequeña.] Por tanto, para $y_1 \leq y_2$, $P(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2)$ es igual a la probabilidad de la unión de los dos eventos $(Y_1 \leq y_1, Y_2 \leq y_2)$ y $(Y_2 \leq y_1, Y_1 \leq y_2)$. Esto es,

$$P(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2) = P[(Y_1 \leq y_1, Y_2 \leq y_2) \cup (Y_2 \leq y_1, Y_1 \leq y_2)].$$

Usando la ley aditiva de la probabilidad y tomando en cuenta que $y_1 \leq y_2$, vemos que

$$\begin{aligned} P(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2) = {} & P(Y_1 \leq y_1, Y_2 \leq y_2) + P(Y_2 \leq y_1, Y_1 \leq y_2) \\ & - P(Y_1 \leq y_1, Y_2 \leq y_1). \end{aligned}$$

Como Y_1 y Y_2 son independientes y $P(Y_i \leq w) = F(w)$, para $i = 1, 2$, se deduce que, para $y_1 \leq y_2$,

$$\begin{aligned} P(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2) &= F(y_1)F(y_2) + F(y_2)F(y_1) - F(y_1)F(y_1) \\ &= 2F(y_1)F(y_2) - [F(y_1)]^2. \end{aligned}$$

Si $y_1 > y_2$ (recuerde que $Y_{(1)} \leq Y_{(2)}$),

$$\begin{aligned} P(Y_{(1)} \leq y_1, Y_{(2)} \leq y_2) &= P(Y_{(1)} \leq y_2, Y_{(2)} \leq y_2) \\ &= P(Y_1 \leq y_2, Y_2 \leq y_2) = [F(y_2)]^2. \end{aligned}$$

Resumiendo, la función de distribución conjunta de $Y_{(1)}$ y $Y_{(2)}$ es

$$F_{Y_{(1)}Y_{(2)}}(y_1, y_2) = \begin{cases} 2F(y_1)F(y_2) - [F(y_1)]^2, & y_1 \leq y_2, \\ [F(y_2)]^2, & y_1 > y_2. \end{cases}$$

Si denotamos con $g_{(1)(2)}(y_1, y_2)$ la densidad conjunta de $Y_{(1)}$ y $Y_{(2)}$, vemos que, al derivar primero con respecto a y_2 y luego con respecto a y_1,

$$g_{(1)(2)}(y_1, y_2) = \begin{cases} 2f(y_1)f(y_2), & y_1 \leq y_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Se puede aplicar el mismo método para determinar la densidad conjunta de $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$, que resulta ser

$$g_{(1)(2)\cdots(n)}(y_1, y_2, \ldots, y_n) = \begin{cases} n!f(y_1)f(y_2), \ldots, f(y_n), & y_1 \leq y_2 \leq \cdots \leq y_n, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La función de densidad marginal para cualquiera de los estadísticos de orden se puede hallar a partir de esta función de densidad conjunta, de lo cual no nos ocuparemos formalmente en este texto.

EJEMPLO **6.16** Componentes electrónicos de cierto tipo tienen una vida útil Y, con densidad de probabilidad dada por

$$f(y) = \begin{cases} (1/100)e^{-y/100}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

(La duración de la vida útil se expresa en horas.) Suponga que dos de estos componentes operan de manera independiente y en serie en cierto sistema (en consecuencia, el sistema falla cuando cualquiera de los dos componentes deja de funcionar). Encuentre la función de densidad para X, la duración de la vida útil del sistema.

Solución Como el sistema deja de funcionar cuando falla el primer componente, $X = \text{mín}(Y_1, Y_2)$, donde Y_1 y Y_2 son variables aleatorias independientes con la densidad dada. Entonces, como $F(y) = 1 - e^{-y/100}$, para $y \geq 0$,

$$f_X(y) = g_{(1)}(y) = n[1 - F(y)]^{n-1} f(y)$$
$$= \begin{cases} 2e^{-y/100}(1/100)e^{-y/100}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y se deduce que

$$f_X(y) = \begin{cases} (1/50)e^{-y/50}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por tanto, el mínimo de las dos variables aleatorias exponencialmente distribuidas tiene una distribución exponencial. Observe que la duración media de cada uno de los componentes es 100 horas, mientras que la duración media del sistema es $E(X) = E(Y_{(1)} = 50 = 100/2$. ■

EJEMPLO **6.17** Suponga que los componentes del Ejemplo 6.16 operan en paralelo (por tanto, el sistema no falla sino hasta que fallen los dos componentes). Encuentre la función de densidad de X, la duración del sistema.

Solución Ahora $X = \text{máx}(Y_1, Y_2)$, y

$$f_X(y) = g_{(2)}(y) = n[F(y)]^{n-1} f(y)$$
$$= \begin{cases} 2(1 - e^{-y/100})(1/100)e^{-y/100}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y, por tanto,

$$f_X(y) = \begin{cases} (1/50)(e^{-y/100} - e^{-y/50}), & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Vemos aquí que el valor máximo de las dos variables aleatorias exponenciales no es una variable aleatoria exponencial. ■

Aun cuando una deducción rigurosa de la función de densidad del estadístico de k-ésimo orden (k es un entero, $1 < k < n$) es un poco complicada, la función de densidad resultante tiene una estructura intuitivamente sensible. Una vez que se comprende dicha estructura, la densidad se puede escribir con facilidad. Suponga que la función de densidad de una variable aleatoria continua en un punto particular es proporcional a la probabilidad de que la variable sea "cercana" a ese punto. Esto es, si Y es una variable aleatoria continua con función de densidad $f(y)$, entonces

$$P(y \leq Y \leq y + dy) \approx f(y)\,dy.$$

Ahora considere el estadístico de k-ésimo orden, $Y(k)$. Si el k-ésimo valor más grande es cercano a y_k, entonces $k - 1$ de los valores de Y debe ser menor que y_k, una de las Y debe ser cercana a y_k, y los restantes $n - k$ valores de las Y deben ser mayores que y_k. Recuerde la distribución multinomial de la Sección 5.9. En este caso, tenemos tres clases de valores de Y:

Clase 1: las Y que tienen valores menores que y_k necesitan $k - 1$.
Clase 2: las Y que tienen valores cercanos que y_k necesitan 1.
Clase 3: las Y que tienen valores mayores que y_k necesitan $n - k$.

Las probabilidades de cada una de estas clases son, respectivamente, $p_1 = P(Y < y_k) = F(y_k)$, $p_2 = P(y_k \leq Y \leq y_k + dy_k) \approx f(y_k)dy_k$, y $p_3 = P(y > y_k) = 1 - F(y_k)$. Usando las probabilidades multinomiales estudiadas antes, vemos que

$$\begin{aligned}
P(y_k \leq Y_{(k)} \leq y_k + dy_k) &\\
&\approx P[(k-1) \text{ de la clase 1, 1 de la clase 2, } (n-k) \text{ de la clase 3}]\\
&\approx \binom{n}{k-1 \quad 1 \quad n-k} p_1^{k-1}\, p_2^1\, p_3^{n-k}\\
&\approx \frac{n!}{(k-1)!\,1!\,(n-k)!}\left\{[F(y_k)]^{k-1}\, f(y_k)\,dy_k\,[1 - F(y_k)]^{n-k}\right\}
\end{aligned}$$

y

$$g_{(k)}(y_k)\,dy_k \approx \frac{n!}{(k-1)!\,1!\,(n-k)!} F^{k-1}(y_k)\, f(y_k)\,[1 - F(y_k)]^{n-k}\,dy_k.$$

La densidad del estadístico de k-ésimo orden y la densidad conjunta de estadísticos de orden dos se dan en el teorema siguiente.

TEOREMA 6.5

Sean $Y_1, \ldots, Y_n$ variables aleatorias continuas distribuidas idénticamente e independientes, con función de distribución común $F(y)$ y función de densidad común $f(y)$. Si $Y_{(k)}$ denota el estadístico de orden k-ésimo, entonces la función de densidad de $Y_{(k)}$ está dada por

$$g_{(k)}(y_k) = \frac{n!}{(k-1)!\,(n-k)!}[F(y_k)]^{k-1}\,[1 - F(y_k)]^{n-k} f(y_k), \qquad -\infty < y_k < \infty.$$

Si j y k son dos enteros tales que $1 \leq j < k \leq n$, la densidad conjunta de $Y_{(j)}$ y $Y_{(k)}$ está dada por

$$\begin{aligned} g_{(j)(k)}(y_j,\ y_k) = {} & \frac{n!}{(j-1)!\,(k-1-j)!\,(n-k)!}[F(y_j)]^{j-1} \\ & \times [F(y_k) - F(y_j)]^{k-1-j} \times [1 - F(y_k)]^{n-k}\, f(y_j)\, f(y_k), \\ & \quad -\infty < y_j < y_k < \infty. \end{aligned}$$

La derivación intuitiva, heurística, de la densidad conjunta dada en el Teorema 6.5 es similar a la establecida antes para la densidad de un estadístico de un solo orden. Para $y_j < y_k$, la densidad conjunta se puede interpretar como la probabilidad de que la j-ésima observación más grande sea cercana a y_j y la k-ésima más grande sea cercana a y_k. Defina las cinco clases de valores de Y:

Clase 1: las Y que tengan valores menores que y_j necesitan $j - 1$.
Clase 2: las Y que tengan valores cercanos a y_j necesitan 1.
Clase 3: las Y que tengan valores entre y_j y y_k necesitan $k - 1 - j$.
Clase 4: las Y que tengan valores cercanos a y_k necesitan 1.
Clase 5: las Y que tengan valores mayores que y_k necesitan $n - k$.

De nuevo, use la distribución multinomial para completar el argumento heurístico.

EJEMPLO **6.18** Suponga que $Y_1, Y_2, \ldots, Y_5$ denota una muestra aleatoria de una distribución uniforme definida en el intervalo (0, 1). Esto es,

$$f(y) = \begin{cases} 1, & 0 \leq y \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad para el estadístico de segundo orden. También, proporcione la función de densidad conjunta para los estadísticos de segundo y cuarto órdenes.

Solución La función de distribución asociada con cada una de las Y es

$$F(y) = \begin{cases} 0, & y < 0, \\ y, & 0 \leq y \leq 1, \\ 1, & y > 1. \end{cases}$$

La función de densidad del estadístico de segundo orden, $Y_{(2)}$, se puede obtener directamente del Teorema 6.5 con $n = 5$, $k = 2$. Así, con $f(y)$ y $F(y)$ como se vio,

$$\begin{aligned} g_{(2)}(y_2) &= \frac{5!}{(2-1)!\,(5-2)!}[F(y_2)]^{2-1}[1 - F(y_2)]^{5-2}\, f(y_2), \quad -\infty < y_2 < \infty, \\ &= \begin{cases} 20y_2(1 - y_2)^3, & 0 \leq y_2 \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases} \end{aligned}$$

La anterior es una densidad beta con $\alpha = 2$ y $\beta = 4$. En general, el estadístico de k-ésimo orden basado en una muestra de tamaño n desde una distribución uniforme (0, 1) tiene una densidad beta con $\alpha = k$ y $\beta = n - k + 1$.

La densidad conjunta para los estadísticos de segundo y cuarto órdenes se obtiene fácilmente del segundo resultado del Teorema 6.5. Con $f(y)$ y $F(y)$ como antes, $j = 2$, $k = 4$ y $n = 5$,

$$\begin{aligned} g_{(2)(4)}(y_2, y_4) &= \frac{5!}{(2-1)!\,(4-1-2)!\,(5-4)!}[F(y_2)]^{2-1}[F(y_4) - F(y_2)]^{4-1-2} \\ &\quad \times [1 - F(y_4)]^{5-4} f(y_2) f(y_4), \qquad -\infty < y_2 < y_4 < \infty \\ &= \begin{cases} 5!\, y_2(y_4 - y_2)(1 - y_4), & 0 \le y_2 < y_4 \le 1 \\ 0, & \text{en cualquier otro punto.} \end{cases} \end{aligned}$$

Desde luego, esta densidad conjunta se puede usar para evaluar probabilidades conjuntas acerca de $Y_{(2)}$ y $Y_{(4)}$ o para evaluar el valor esperado de funciones de estas dos variables. ■

Ejercicios

6.72 Sean Y_1 y Y_2 independientes y distribuidas uniformemente en el intervalo (0, 1). Encuentre

a la función de densidad de probabilidad de $U_1 = \text{mín}(Y_1, Y_2)$.

b $E(U_1)$ y $V(U_1)$.

6.73 Al igual que en el Ejercicio 6.72, sean Y_1 y Y_2 independientes y distribuidas uniformemente en el intervalo (0, 1). Encuentre

a la función de densidad de probabilidad de $U_2 = \text{máx}(Y_1, Y_2)$,

b $E(U_2)$ y $V(U_2)$,

6.74 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes distribuidas uniformemente en el intervalo $[0, \theta]$. Encuentre la

a función de distribución de probabilidad de $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$,

b función de densidad de $Y_{(n)}$,

c media y varianza de $Y_{(n)}$.

6.75 Consulte el Ejercicio 6.74. Suponga que el número de minutos que espera para abordar un autobús está distribuido uniformemente en el intervalo [0, 15]. Si una persona toma el autobús cinco veces, ¿cuál es la probabilidad de que su espera más larga sea menor que 10 minutos?

***6.76** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias distribuidas uniformemente e independientes en el intervalo $[0, \theta]$.

a Encuentre la función de densidad de $Y_{(k)}$, el estadístico de k-ésimo orden, donde k es un entero entre 1 y n.

b Use el resultado del inciso a para determinar $E(Y_{(k)})$.

c Encuentre $V(Y_{(k)})$.

d Utilice el resultado del inciso c para hallar $E(Y_{(k)} - Y_{(k-1)})$, la diferencia media entre dos estadísticos de orden sucesivo. Interprete este resultado.

***6.77** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias distribuidas uniformemente e independientes en el intervalo $[0, \theta]$.

a Encuentre la función de densidad conjunta de $Y_{(j)}$ y $Y_{(k)}$ donde j y k son enteros $1 \le j < k \le n$.

b Utilice el resultado del inciso a para hallar $\text{Cov}(Y_{(j)}, Y_{(k)})$ cuando j y k son enteros $1 \le j < k \le n$.

c Utilice el resultado del inciso b y el Ejercicio 6.76 para hallar $V(Y_{(k)} - Y_{(j)})$, la varianza de la diferencia entre estadísticos de orden dos.

6.78 Consulte el Ejercicio 6.76. Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias uniformemente distribuidas e independientes en el intervalo [0, 1], demuestre que $Y_{(k)}$, la estadística de k-ésimo orden, tiene una función de densidad beta con $\alpha = k$ y $\beta = n - k + 1$.

6.79 Consulte el Ejercicio 6.77. Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes uniformemente distribuidas en el intervalo $[0, \theta]$, demuestre que $U = Y_{(1)}/Y_{(n)}$ y $Y_{(n)}$ son independientes.

6.80 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes, cada una con una distribución beta, con $\alpha = \beta = 2$. Encuentre

a la función de distribución de probabilidad de $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$,
b la función de densidad de $Y_{(n)}$,
c $E(Y_{(n)})$ cuando $n = 2$.

6.81 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias exponencialmente distribuidas e independientes con media β.

a Demuestre que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ tiene una distribución exponencial, con media β/n.
b Si $n = 5$ y $\beta = 2$, encuentre $P(Y_{(1)} \leq 3.6)$.

6.82 Si Y es una variable aleatoria continua y m es la mediana de la distribución, entonces m es tal que $P(Y \leq m) = P(Y \geq m) = 1/2$. Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias exponencialmente distribuidas e independientes con media β y mediana m, el Ejemplo 6.17 implica que $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ *no* tiene una distribución exponencial. Use la forma general de $F_{Y_{(n)}}(y)$ para demostrar que $P(Y_{(n)} > m) = 1 - (.5)^n$.

6.83 Consulte el Ejercicio 6.82. Si $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de cualquier distribución continua con media m, ¿cuál es $P(Y_{(n)} > m)$?

6.84 Consulte el Ejercicio 6.26. La función de densidad Weibull está dada por

$$f(y) = \begin{cases} \dfrac{1}{\alpha} m y^{m-1} e^{-y^m/\alpha}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde α y m son constantes positivas. Si una muestra aleatoria de tamaño n se toma de una población con distribución Weibull, encuentre la función de distribución y la función de densidad para $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$. ¿$Y_{(1)}$ tiene una distribución Weibull?

6.85 Sean Y_1 y Y_2 independientes y uniformemente distribuidas en el intervalo (0, 1). Encuentre $P(2Y_{(1)} < Y_{(2)})$.

***6.86** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes exponencialmente distribuidas con media β. Obtenga la

a función de densidad para $Y_{(k)}$, el estadístico de k-ésimo orden, donde k es un entero entre 1 y n.
b función de densidad conjunta para $Y_{(j)}$ y $Y_{(k)}$ donde j y k son enteros $1 \leq j < k \leq n$.

6.87 Los precios de apertura por acción Y_1 y Y_2 de dos acciones similares son variables aleatorias independientes, cada una con una función de densidad dada por

$$f(y) = \begin{cases} (1/2)e^{-(1/2)(y-4)}, & y \geq 4, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En una mañana determinada, un inversionista va a comprar acciones de la emisión menos costosa. Encuentre

a la función de densidad de probabilidad para el precio por acción que el inversionista pagará.

b el costo esperado por acción que el inversionista pagará.

6.88 Suponga que el tiempo Y que un trabajador tarda en completar cierta tarea tiene la función de densidad de probabilidad dada por

$$f(y) = \begin{cases} e^{-(y-\theta)}, & y > \theta, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde θ es una constante positiva que representa el tiempo mínimo hasta la terminación del trabajo. Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tiempos de terminación de esta distribución. Encuentre

a la función de densidad para $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$.

b $E(Y_{(1)})$.

***6.89** Si $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria de la distribución uniforme $f(y) = 1$, $0 \leq y \leq 1$ encuentre la función de densidad de probabilidad de la amplitud $R = Y_{(n)} - Y_{(1)}$.

***6.90** Suponga que el número de veces que ocurre cierto evento en el intervalo $(0, t)$ tiene una distribución de Poisson. Si sabemos que n de tales eventos han ocurrido en $(0, t)$, entonces los tiempos reales, medidos desde 0, para los sucesos del evento en cuestión forman un conjunto ordenado de variables aleatorias, que denotamos con $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n)}$. [$W_{(i)}$ en realidad es el tiempo de espera de 0 hasta que ocurre el i-ésimo evento.] Se puede demostrar que la función de densidad conjunta para $W_{(1)}, W_{(2)}, \ldots, W_{(n)}$ está dada por

$$f(w_1, w_2, \ldots, w_n) = \begin{cases} \dfrac{n!}{t^n}, & w_1 \leq w_2 \leq \cdots \leq w_n, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

[Esta es la función de densidad para la muestra ordenada de tamaño n desde una distribución uniforme en el intervalo $(0, t)$.] Suponga que las llamadas telefónicas que entran en un conmutador siguen una distribución de Poisson con una media de diez llamadas por minuto. En un periodo lento de dos minutos de duración sólo entraron cuatro llamadas. Encuentre

a la probabilidad de que las cuatro llamadas entren durante el primer minuto; es decir, encuentre $P(W_{(4)} \leq 1)$,

b el tiempo de espera desde el principio de los dos minutos hasta la cuarta llamada.

***6.91** Suponga que n componentes electrónicos, cada uno con una vida útil distribuida exponencialmente con media θ, se ponen en operación al mismo tiempo. Los componentes operan en forma independiente y se observan hasta que r de ellos fallan ($r \leq n$). Sea W_j el tiempo que transcurre hasta que el componente j-ésimo falla, con $W_1 \leq W_2 \leq \cdots \leq W_r$. Sea $T_j = W_j - W_{j-1}$ para $j \geq 2$ y $T_1 = W_1$. Observe que T_j mide el tiempo transcurrido entre fallas sucesivas.

a Demuestre que T_j, para $j = 1, 2, \ldots, r$, tiene una distribución exponencial con media $\theta/(n - j + 1)$.

b Demuestre que

$$U_r = \sum_{j=1}^{r} W_j + (n - r)W_r = \sum_{j=1}^{r}(n - j + 1)T_j$$

y, en consecuencia, que $E(U_r) = r\theta$. [U_r se denomina *vida útil total observada* y podemos usar U_r/r como una aproximación (o "estimador") de θ.]

6.8 Resumen

En este capítulo abordamos las distribuciones de probabilidad para funciones de variables aleatorias. Este es un problema importante en estadística porque los estimadores de parámetros poblacionales son funciones de variables aleatorias. Por consiguiente, es necesario conocer un poco acerca de las distribuciones de probabilidad de estas funciones, o estimadores, para evaluar la bondad de nuestros procedimientos estadísticos. Un análisis de la estimación se presenta en los Capítulos 8 y 9.

Los métodos para determinar las distribuciones de probabilidad para funciones de variables aleatorias son el de las funciones de distribución (Sección 6.3), el de las transformaciones (Sección 6.4) y el de las funciones generadoras de momentos (Sección 6.5). Debe observarse que no hay un particular mejor para todas las situaciones, ya que la solución depende en gran medida de la naturaleza de la función de que se trate. Si U_1 y U_2 son dos funciones de las variables aleatorias continuas Y_1 y Y_2, la función de densidad *conjunta* para U_1 y U_2 se puede determinar usando la técnica del jacobiano de la Sección 6.6. La facilidad para manejar estos métodos se puede obtener sólo por medio de la práctica. Los ejercicios al final de cada sección y de cada capítulo son un buen punto de partida.

Las funciones de densidad de estadísticos de orden se presentaron en la Sección 6.7.

En el capítulo 7 se considerarán algunas funciones especiales de variables aleatorias que son particularmente útiles en la inferencia estadística.

Bibliografía y lecturas adicionales

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

Hoel, P. G. 1984. *Introductionto Mathematical Statistics*, 5th ed. New York: Wiley.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3ded. New York: McGraw-Hill.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Ejercicios complementarios

6.92 Si Y_1 y Y_2 son variables aleatorias normales independientes y distribuidas idénticamente con media μ y varianza σ^2, encuentre la función de densidad de probabilidad para $U = (1/2)(Y_1 - 3Y_2)$.

6.93 Cuando la corriente I pasa por la resistencia R, la potencia generada está dada por $W = I^2R$. Suponga que I tiene una distribución uniforme en el intervalo (0, 1) y R tiene una función de densidad dada por

$$f(r) = \begin{cases} 2r, & 0 \le r \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad de probabilidad para W. (Suponga que I es independiente de R.)

6.94 Dos expertos en eficiencia toman medidas independientes Y_1 y Y_2 acerca del tiempo que tardan unos trabajadores para completar cierta tarea. Se supone que cada una de las medidas tiene una función de densidad dada por

$$f(y) = \begin{cases} (1/4)ye^{-y/2}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad para el promedio $U = (1/2)(Y_1 + Y_2)$. [*Sugerencia:* use el método de funciones generadoras de momento.]

6.95 Sean Y_1 y Y_2 independientes y uniformemente distribuidas en el intervalo (0, 1). Encuentre la función de densidad de probabilidad de cada uno de los siguientes:

a $U_1 = Y_1/Y_2$.

b $U_2 = -\ln(Y_1 Y_2)$.

c $U_3 = Y_1 Y_2$.

6.96 Suponga que Y_1 está normalmente distribuida con media 5 y varianza 1 y Y_2 está normalmente distribuida con media 4 y varianza 3. Si Y_1 y Y_2 son independientes, ¿cuál es $P(Y_1 > Y_2)$?

***6.97** Suponga que Y_1 es una variable aleatoria binomial con cuatro intentos y probabilidad de éxito .2 y que Y_2 es una variable aleatoria binomial independiente con tres intentos y probabilidad de éxito .5. Sea $W = Y_1 + Y_2$. De acuerdo con el Ejercicio 6.53(e), W *no* tiene distribución binomial. Encuentre la función de masa de probabilidad para W. [*Sugerencia:* $P(W = 0) = P(Y_1 = 0, Y_2 = 0)$; $P(W = 1) = P(Y_1 = 1, Y_2 = 0) + P(Y_1 = 0, Y_2 = 1)$; etc.]

6.98 El tiempo que una máquina opera sin falla está denotado por Y_1 y el tiempo para reparar una falla, por Y_2. Después de hacer una reparación, se supone que la máquina opera como nueva. Y_1 y Y_2 son independientes y cada uno tiene la función de densidad

$$f(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad de probabilidad para $U = Y_1/(Y_1 + Y_2)$, la proporción de tiempo en que la máquina está en operación durante cualquier ciclo de operación-reparación.

***6.99** Consulte el Ejercicio 6.98. Demuestre que U, la proporción del tiempo que la máquina está operando durante cualquier ciclo de operación-reparación, es independiente de $Y_1 + Y_2$, la duración del ciclo.

6.100 El tiempo hasta que se presenta una falla en un aparato electrónico tiene una distribución exponencial con media de 15 meses. Si se prueba una muestra aleatoria de cinco de esos aparatos, ¿cuál es la probabilidad de que la primera falla entre los cinco aparatos ocurra

a después de 9 meses?,

b antes de 12 meses?

***6.101** Una paracaidista desea caer en un blanco T, pero encuentra que es igualmente probable que caiga en cualquier punto sobre una recta (A, B), en la que T está en el punto medio. Encuentre la función de densidad de probabilidad de la distancia entre el punto de caída de la paracaidista y el blanco. [*Sugerencia:* denote -1 con A, $+1$ con B y 0 con T. Entonces el punto de caída de la paracaidista tiene una coordenada X, que está distribuida uniformemente entre -1 y $+1$. La distancia entre X y T es $|X|$.]

6.102 Dos policías son enviados a patrullar un camino de 1 milla de largo. Los policías son asignados a puntos escogidos independientemente y al azar a lo largo del camino. Encuentre la probabilidad de que los policías estén a menos de 1/2 milla entre sí cuando lleguen a sus puestos asignados.

***6.103** Sean Y_1 y Y_2 variables aleatorias normales estándar e independientes. Encuentre la función de densidad de probabilidad de $U = Y_1/Y_2$.

6.104 Sean Y_1 y Y_2 variables aleatorias independientes, cada una con la misma distribución geométrica.

a Encuentre $P(Y_1 = Y_2) = P(Y_1 - Y_2 = 0)$. [*Sugerencia:* su respuesta debe incluir la evaluación de una serie geométrica infinita. Los resultados del Apéndice A1.11 le serán útiles.]

b Encuentre $P(Y_1 - Y_2 = 1)$.

***c** Si $U = Y_1 - Y_2$, encuentre la función de probabilidad (discreta) para U. [*Sugerencia:* el inciso a da $P(U = 0)$ y el inciso b da $P(U = 1)$. Considere por separado los valores enteros positivos y negativos para U.]

6.105 Una variable aleatoria Y tiene una *distribución beta de segunda clase*, si, para $\alpha > 0$ y $\beta > 0$, su densidad es

$$f_Y(y) = \begin{cases} \dfrac{y^{\alpha-1}}{B(\alpha,\beta)(1+y)^{\alpha+\beta}}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

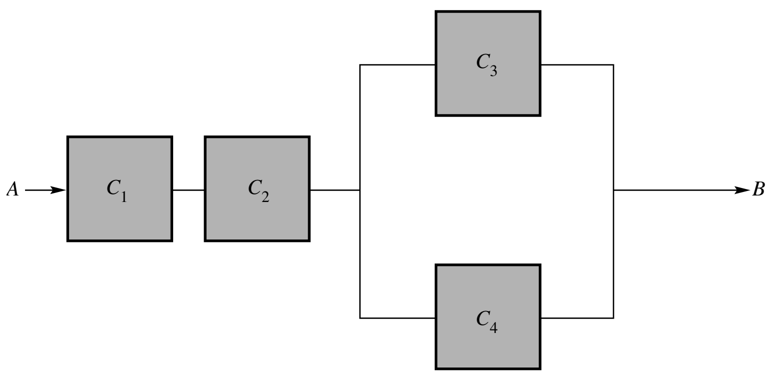
Deduzca la función de densidad de $U = 1/(1 + Y)$.

6.106 Si Y es una variable aleatoria continua con función de distribución $F(y)$, encuentre la función de densidad de probabilidad de $U = F(Y)$.

6.107 Sea Y distribuida uniformemente en el intervalo $(-1, 3)$. Encuentre la función de densidad de probabilidad de $U = Y^2$.

6.108 Si Y denota la vida útil de un componente y $F(y)$ es la función de distribución de Y, entonces $P(Y > y) = 1 - F(y)$ se llama *confiabilidad* del componente. Suponga que un sistema formado por cuatro componentes con funciones de confiabilidad idénticas, $1 - F(y)$, opera como se indica en la Figura 6.10. El sistema opera correctamente si una cadena ininterrumpida de componentes está en operación entre A y B. Si los cuatro componentes operan independientemente, encuentre la confiabilidad del sistema en términos de $F(y)$.

FIGURA **6.10** Diagrama de circuito



6.109 El porcentaje de alcohol en cierto compuesto es una variable aleatoria Y, con la siguiente función de densidad:

$$f(y) = \begin{cases} 20y^3(1-y), & 0 < y < 1 \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Suponga que el precio de venta del compuesto depende de su contenido de alcohol. Específicamente, si $1/3 < y < 2/3$, el compuesto se vende en C_1 dólares por galón; en cualquier otro punto, se vende en C_2 dólares por galón. Si el costo de producción es C_3 dólares por galón, encuentre la distribución de probabilidad de la utilidad por galón.

6.110 Un ingeniero ha observado que los tiempos de separación entre vehículos que pasan por cierto punto en una carretera, tienen una distribución exponencial con media de 10 segundos. Encuentre la

a probabilidad de que el siguiente tiempo de separación observado no sea de más de un minuto.

b probabilidad de la función de densidad para la suma de los siguientes cuatro tiempos de separación que van a ser observados. ¿Qué suposiciones son necesarias para que esta respuesta sea correcta?

***6.111** Si una variable aleatoria U está distribuida normalmente con media μ y varianza σ^2 y $Y = e^U$ [o bien, lo que es equivalente, $U = \ln(Y)$], entonces se dice que Y tiene una *distribución log-normal*. La distribución log-normal se usa con frecuencia en ciencias biológicas y físicas para modelar magnitudes, ya sea de volumen o peso, de diversas cantidades, como pueden ser partículas de carbón, colonias de bacterias y animales individuales. Sean U y Y las variables establecidas. Demuestre que

a la función de densidad para Y es

$$f(y) = \begin{cases} \left(\dfrac{1}{y\sigma\sqrt{2\pi}}\right) e^{-(\ln y-\mu)^2/(2\sigma^2)}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b $E(Y) = e^{\mu+(\sigma^2/2)}$ y $V(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. [*Sugerencia:* recuerde que $E(Y) = E(e^U)$ y $E(Y^2) = E(e^{2U})$, donde U está distribuida normalmente con media μ y varianza σ^2. Recuerde que la función generadora de momentos de U es $m_U(t) = e^{tU}$.]

***6.112** Si una variable aleatoria U tiene una distribución gamma con parámetros $\alpha > 0$ y $\beta > 0$, entonces se dice que $Y = e^U$ [o bien, lo que es lo mismo, $U = \ln(Y)$] tiene una *distribución log-gamma.* La distribución log-gamma es utilizada por actuarios como parte de un importante modelo para la distribución de reclamaciones de seguros. Sean U y Y las variables establecidas.

a Demuestre que la función de densidad para Y es

$$f(y) = \begin{cases} \left[\dfrac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{-(1+\beta)/\beta}(\ln y)^{\alpha-1}, & y > 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b Si $\beta < 1$, demuestre que $E(Y) = (1 - \beta)^{-\alpha}$. [Vea la sugerencia para el inciso c.]

c Si $\beta < .5$, demuestre que $V(Y) = (1 - 2\beta)^{-\alpha} - (1 - \beta)^{-2\alpha}$. [*Sugerencia:* recuerde que $E(Y) = E(e^U)$ y $E(Y^2) = E(e^{2U})$, donde U tiene distribución gamma con parámetros $\alpha > 0$ y $\beta > 0$, y que la función generadora de momento de una variable aleatoria con distribución gamma sólo existe si $t < \beta^{-1}$; vea el Ejemplo 4.13.]

***6.113** Considere que (Y_1, Y_2) tiene función de densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ y sean $U_1 = Y_1Y_2$ y $U_2 = Y_2$.

a Demuestre que la densidad conjunta de (U_1, U_2) es

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}\left(\frac{u_1}{u_2}, u_2\right)\frac{1}{|u_2|}.$$

b Demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}\left(\frac{u_1}{u_2}, u_2\right)\frac{1}{|u_2|}\,du_2.$$

c Si Y_1 y Y_2 son independientes, demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1}\left(\frac{u_1}{u_2}\right) f_{Y_2}(u_2)\frac{1}{|u_2|}\,du_2.$$

***6.114** Una máquina produce recipientes esféricos cuyos radios varían de acuerdo con la función de densidad de probabilidad dada por

$$f(r) = \begin{cases} 2r, & 0 \leq r \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la función de densidad de probabilidad para el volumen de los recipientes.

***6.115** Denote con v el volumen de una figura tridimensional. Sea Y el número de partículas observadas en el volumen v y suponga que Y tiene una distribución de Poisson con media λv. Las partículas podrían representar contaminantes del aire, bacterias en agua o estrellas en el cielo.

a Si un punto se escoge al azar dentro del volumen v, demuestre que la distancia R a la partícula más cercana tiene la función de densidad de probabilidad dada por

$$f(r) = \begin{cases} 4\lambda\pi r^2 e^{-(4/3)\lambda\pi r^3}, & r > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b Si R es como en el inciso a, demuestre que $U = R^3$ tiene una distribución exponencial.

***6.116** Sea (Y_1, Y_2) que tiene función de densidad conjunta $f_{Y_1,Y_2}(y_1, y_2)$ y sean $U_1 = Y_1 - Y_2$ y $U_2 = Y_2$.

a Demuestre que la densidad conjunta de (U_1, U_2) es

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}(u_1 + u_2, u_2).$$

b Demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(u_1 + u_2, u_2)\, du_2.$$

c Si Y_1 y Y_2 son independientes, demuestre que la función de densidad marginal para U_1 es

$$f_{U_1}(u_1) = \int_{-\infty}^{\infty} f_{Y_1}(u_1 + u_2) f_{Y_2}(u_2)\, du_2.$$

CAPÍTULO 7

Distribuciones muestrales y el teorema del límite central

7.1 Introducción

En el Capítulo 6 presentamos métodos para hallar las distribuciones de funciones de variables aleatorias. A lo largo de este capítulo trabajaremos con funciones de las variables $Y_1, Y_2, \ldots, Y_n$ observadas en una muestra aleatoria seleccionada de una población de interés. Como se explicó en el Capítulo 6, las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ son independientes y tienen la misma distribución. Algunas funciones de las variables aleatorias observadas en una muestra se usan para calcular o tomar decisiones acerca de parámetros desconocidos de la población.

Por ejemplo, suponga que deseamos estimar una media poblacional μ. Si obtenemos una muestra aleatoria de n observaciones, $y_1, y_2, \ldots, y_n$, parece razonable estimar μ con la media muestral

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

La bondad de esta estimación depende del comportamiento de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ y el efecto que este comportamiento tiene sobre $\overline{Y} = (1/n)\sum_{i=1}^{n} Y_i$. Observe que la variable aleatoria $\overline{Y}$ es una función de (sólo) las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ y el tamaño muestral n (constante). La variable aleatoria $\overline{Y}$ es por tanto un ejemplo de un *estadístico*.

DEFINICIÓN 7.1

Un *estadístico* es una función de las variables aleatorias observables en una muestra y de constantes conocidas.

Usted ha encontrado numerosas estadísticas, la media muestral $\overline{Y}$, la varianza muestral S^2, $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$, $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$, la amplitud $R = Y_{(n)} - Y_{(1)}$, la mediana muestral, etcétera. Se usan estadísticos para hacer inferencias (estimaciones o decisiones) acerca de parámetros de población desconocidos. Como todos los estadísticos son funciones de las variables aleatorias observadas en una muestra, también son variables aleatorias. En consecuencia, todos los estadísticos tienen distribuciones de probabilidad, que llamaremos sus *distribuciones muestrales*. Desde un punto de vista práctico, la distribución muestral de un estadístico proporciona un modelo teórico para el histograma de frecuencia relativa de los posibles valores del estadístico que observaríamos por medio de muestreo repetido.

El siguiente ejemplo contiene una distribución de muestreo de la media muestral cuando se obtienen muestras de una población conocida asociada con lanzar al aire un dado sin cargar.

EJEMPLO 7.1 Un dado sin cargar se lanza tres veces. Sean Y_1, Y_2 y Y_3 el número de puntos vistos en la cara superior para los tiros 1, 2 y 3, respectivamente. Suponga que estamos interesados en $\overline{Y} = (Y_1 + Y_2 + Y_3)/3$, el número promedio de puntos vistos en una muestra de tamaño 3. ¿Cuáles son la media $\mu_{\overline{Y}}$ y la desviación estándar $\sigma_{\overline{Y}}$, de $\overline{Y}$? ¿Cómo podemos determinar la distribución muestral de $\overline{Y}$?

Solución En el Ejercicio 3.22 se demostró que $\mu = E(Y_i) = 3.5$ y $\sigma^2 = V(Y_i) = 2.9167$, $i = 1, 2, 3$. Como Y_1, Y_2 y Y_3 son variables aleatorias independientes, el resultado obtenido en el Ejemplo 5.27 (usando el Teorema 5.12) implica que

$$E(\overline{Y}) = \mu = 3.5, \qquad V(\overline{Y}) = \frac{\sigma^2}{3} = \frac{2.9167}{3} = .9722, \qquad \sigma_{\overline{Y}} = \sqrt{.9722} = .9860.$$

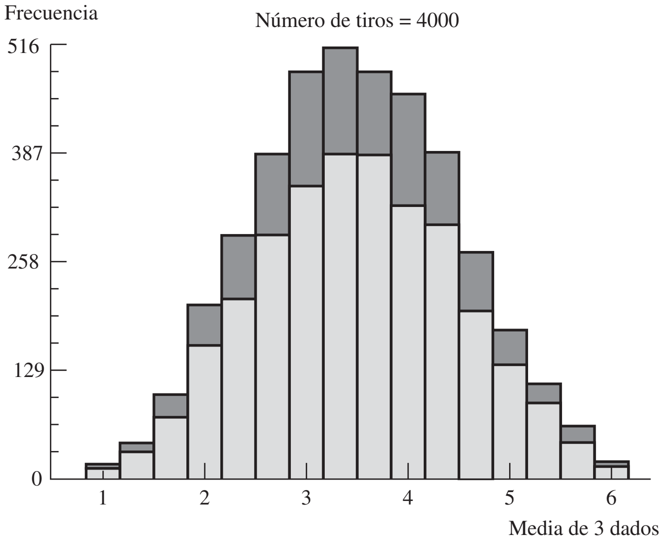
¿Cómo podemos deducir la distribución de la variable aleatoria $\overline{Y}$? Los posibles valores de la variable aleatoria $W = Y_1 + Y_2 + Y_3$ son 3, 4, 5, . . . , 18 y $\overline{Y} = W/3$. Como el dado está equilibrado, es decir, no cargado, cada uno de los $6^3 = 216$ valores distintos de la variable aleatoria multivariante (Y_1, Y_2, Y_3) son igualmente probables y

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = p(y_1, y_2, y_3) = 1/216,$$
$$y_i = 1, 2, \ldots, 6, \quad i = 1, 2, 3.$$

Por tanto,

$$\begin{aligned} P(\overline{Y} = 1) &= P(W = 3) = p(1, 1, 1) = 1/216 \\ P(\overline{Y} = 4/3) &= P(W = 4) = p(1, 1, 2) + p(1, 2, 1) + p(2, 1, 1) = 3/216 \\ P(\overline{Y} = 5/3) &= P(W = 5) = p(1, 1, 3) + p(1, 3, 1) + p(3, 1, 1) \\ &\qquad + p(1, 2, 2) + p(2, 1, 2) + p(2, 2, 1) = 6/216. \end{aligned}$$

Las probabilidades $P(\overline{Y} = i/3)$, $i = 7, 8, \ldots, 18$ se obtienen de manera análoga. ■



(a)

Prob. de pobl. (1) 0.167 (2) 0.167 (3) 0.167 (4) 0.167 (5) 0.167 (6) 0.167
Población: Media = 3.500 Desv.Est. = 1.708
Muestras = 4000 de tamaño 3
Media = 3.495
Desv. Est. = 0.981
+/− 1 Desv. Est.: 0.683
+/− 2 Desv. Est.: 0.962
+/− 3 Desv. Est.: 1.000

(b)

FIGURA **7.1** (a) Distribución de muestreo simulado para $\overline{Y}$, Ejemplo 7.1; (b) media y desviación estándar de los 4000 valores simulados de $\overline{Y}$

La deducción de la distribución de muestreo de la variable aleatoria $\overline{Y}$ trazada en el Ejemplo 7.1 utiliza el método de punto de muestra que se introdujo en el Capítulo 2. Aun cuando no es difícil completar los cálculos del Ejemplo 7.1 y dar la distribución de muestreo exacta para $\overline{Y}$, el proceso es tedioso. ¿Cómo podemos tener una idea de la forma de esta distribución de muestro sin molestarnos en completar estos cálculos? Una forma es simular la distribución de muestreo al tomar muestras independientes repetidas, cada una de tamaño 3, calculando el valor observado $\overline{y}$ para cada muestra y construyendo un histograma de estos valores observados. El resultado de una de estas simulaciones se ilustra en la Figura 7.1(a), que es una gráfica obtenida usando la aplicación breve *DiceSample* (disponible en www.thomsonedu.com/statistics/wackerly).

¿Qué puede observar en la Figura 7.1(a)? Como ya dijimos, el máximo valor observado de $\overline{Y}$ es 6 y el valor mínimo es 1. También, los valores obtenidos en la simulación se acumulan en forma de montículo aproximadamente centrado en 3.5, que es la media teórica de $\overline{Y}$. En la Figura 7.1(b) vemos que el promedio y desviación estándar de los 4000 valores simulados de $\overline{Y}$ son muy cercanos a los valores teóricos obtenidos en el Ejemplo 7.1.

Algunos de los ejercicios del final de esta sección utilizan la aplicación breve *DiceSample* para explorar la distribución muestral simulada de $\overline{Y}$ para diferentes tamaños muestrales y para tiros de dados en los que se usan dados “cargados”. Otras aplicaciones se usan para simular las distribuciones muestrales para la media y la varianza de muestras tomadas de una distribución en forma de campana.

Al igual que las distribuciones muestrales simuladas que usted observará en los ejercicios, la forma de la distribución muestral teórica de cualquier estadístico dependerá de la distribución de las variables aleatorias observables de la muestra. En la siguiente sección usaremos los métodos del Capítulo 6 para deducir las distribuciones muestrales para algunos estadísticos empleados para hacer inferencias acerca de los parámetros de una distribución normal.

Ejercicios

7.1 **Ejercicio Applet** En el Ejemplo 7.1 obtuvimos la media y varianza de la variable aleatoria $\overline{Y}$ con base en una muestra de tamaño 3 tomada de una población conocida, la asociada con lanzar al aire un dado balanceado. Recuerde que si Y denota el número de puntos observados en la cara superior en un solo tiro de un dado balanceado, como en el Ejercicio 3.22,

$$\begin{aligned} P(Y = i) &= 1/6, \qquad i = 1, 2, \ldots, 6, \\ \mu = E(Y) &= 3.5, \\ \text{Var}(Y) &= 2.9167. \end{aligned}$$

Use la aplicación *DiceSample* (en www.thomsonedu.com/statistics/wackerly) para completar lo siguiente:

a Use el botón “Roll One Set” para tomar una muestra de tamaño 3 de la población de tiros de dados. ¿Qué valor se obtuvo para la media de esta muestra? ¿Dónde cae este valor en el histograma? ¿El valor obtenido es igual a uno de los posibles valores asociados con un solo tiro de un dado balanceado? ¿Por qué sí o por qué no?

b Use el botón “Roll One Set” para obtener de nuevo otra muestra de tamaño 3 de una población de tiros de dados. ¿Qué valor se obtuvo para la media de esta nueva muestra? El valor obtenido ¿es igual al valor obtenido en el inciso a? ¿Por qué sí o por qué no?

c Use el botón “Roll One Set” ocho veces más para obtener un total de diez valores de la media muestral. Vea el histograma de estas diez medias. ¿Qué se observa? ¿Cuántos valores diferentes para la media muestral se obtuvieron? ¿Qué valores se observaron más de una vez?

d Use el botón “Roll 10 Sets” hasta obtener o graficar 100 valores realizados para la media muestral, $\overline{Y}$. ¿Qué puede observar acerca de la forma del histograma de los 100 valores recabados? Haga clic en el botón “Show Stats” para ver la media y la desviación estándar de los 100 valores $(\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{100})$ que se observaron. ¿Cómo se compara el promedio de los 100 valores de $\overline{y}_i$, $i = 1, 2, \ldots, 100$ con *E*(*Y*), el número esperado de puntos en un solo tiro de un dado balanceado? (Observe que la media y la desviación estándar de Y que usted calculó en el Ejercicio 3.22 se dan en la segunda línea de la pantalla de selección “Stat Report”.)

e ¿Cómo se compara la desviación estándar de los 100 valores de $\overline{y}_i$, $i = 1, 2, \ldots, 100$ con la desviación estándar de Y dada en la segunda línea de la pantalla de selección “Stat Report”?

f Haga clic en el botón “Roll 1000 Sets” unas cuantas veces, observando cambios en el histograma a medida que genere más y más valores de la media muestral. ¿Cómo se compara el histograma resultante con la gráfica dada en la Figura 7.1(a)?

7.2 Consulte el Ejemplo 7.1 y el Ejercicio 7.1.

a Use el método del Ejemplo 7.1 para hallar el valor exacto de $P(\overline{Y} = 2)$.

b Consulte el histograma obtenido en el Ejercicio 7.1(d). ¿Cómo se compara la frecuencia *relativa* que usted observó $\overline{Y} = 2$, con su respuesta al inciso a?

c Si usted fuera a generar 10,000 valores de $\overline{Y}$, ¿qué espera obtener para la frecuencia relativa de observar $\overline{Y} = 2$?

7.3 **Ejercicio Applet** Consulte el Ejercicio 7.1. Use la aplicación *DiceSample* y arrastre hacia abajo a la siguiente parte de la pantalla que corresponde a tomar muestras de tamaño $n = 12$ de la población correspondiente a lanzar un dado balanceado.

a Tome una sola muestra de tamaño $n = 12$ al hacer clic en el botón "Roll One Set". Use el botón "Roll One Set" para generar nueve valores más de la media muestral. ¿Cómo se compara el histograma de valores observados de la media muestral con el histograma observado en el Ejercicio 7.1(c) que estuvo basado en diez muestras cada una de tamaño 3?

b Use el botón "Roll 10 Sets" nueve veces más hasta obtener una gráfica de 100 valores (cada uno basado en una muestra de tamaño $n = 12$) para la media muestral $\overline{Y}$. Haga clic en el botón "Show Stats" para ver la media y la desviación estándar de los 100 valores $(\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{100})$ que observó.

i ¿Cómo se compara el promedio de estos 100 valores de $\overline{y}_i$, $i = 1, 2, \ldots, 100$ con el promedio de los 100 valores (con base en muestras de tamaño $n = 3$) que obtuvo en el Ejercicio 7.1(d)?

ii Divida la desviación estándar de los 100 valores de $\overline{y}_i$, $i = 1, 2, \ldots, 100$ con base en muestras de tamaño 12 que acaba de obtener por la desviación estándar de los 100 valores (con base en muestras de tamaño $n = 3$) que obtuvo en el Ejercicio 7.1. ¿Por qué espera obtener un valor cercano a $1/2$? [*Sugerencia:* $V(\overline{Y}) = \sigma^2/n$.]

c Haga clic en el botón "Toggle Normal". La función de densidad continua (verde) graficada sobre el histograma es la de una variable aleatoria normal con media y desviación estándar igual a la media y desviación estándar de los 100 valores $(\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{100})$ graficados en el histograma. ¿Esta distribución normal parece estar razonablemente aproximada a la distribución descrita por el histograma?

7.4 **Ejercicio Applet** La población correspondiente a la cara superior de un solo tiro de dado balanceado es tal que los seis valores posibles son igualmente probables. ¿Se observarían resultados análogos a los obtenidos en los Ejercicios 7.1 y 7.2 si el dado no estuviera balanceado? Obtenga acceso a la aplicación *DiceSample* y arrastre hacia abajo a la parte de la pantalla que se refiere a "Loaded Die."

a Si el dado está cargado, los seis resultados posibles no son igualmente probables. ¿Cuáles son las probabilidades asociadas con cada resultado? Haga clic en los botones "1 roll", "10 rolls", y/o "1000 rolls" hasta tener una buena idea de las probabilidades asociadas con los valores 1, 2, 3, 4, 5 y 6. ¿Cuál es la forma general del histograma que obtuvo?

b Haga clic en el botón "Show Stats" para ver los verdaderos valores de las probabilidades de los seis valores posibles. Si Y es la variable aleatoria que denota el número de puntos en la cara superior, ¿cuál es el valor para $\mu = E(Y)$? ¿Cuál es el valor de σ, la desviación estándar de Y? [*Sugerencia:* estos valores aparecen en la pantalla "Stat Report".]

c ¿Cuántas veces simuló usted tirar el dado en el inciso a? ¿Cómo se comparan la media y la desviación estándar de los valores simulados con los verdaderos valores $\mu = E(Y)$ y σ? Simule 2000 tiros más y conteste la misma pregunta.

d Arrastre hacia la parte de la pantalla marcada "Rolling 3 Loaded Dice". Haga clic en el botón "Roll 1000 Sets" hasta haber generado 3000 valores observados para la variable aleatoria $\overline{Y}$.

i ¿Cuál es la forma general de la distribución muestral simulada que obtuvo?

ii ¿Cómo se compara la media de los 3000 valores $\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{3000}$ con el valor de $\mu = E(Y)$ calculada en el inciso a? ¿Cómo se compara la desviación estándar de los 3000 valores con $\sigma/\sqrt{3}$?

e Arrastre a la parte de la pantalla marcada "Rolling 12 Loaded Dice".

i En el inciso ii, usted usará la aplicación breve para generar 3000 muestras de tamaño 12, calculará la media de cada muestra observada y graficará estas medias en un histograma. Antes de usar la aplicación, pronostique el valor aproximado que obtendrá para la media y desviación estándar de los 3000 valores de $\overline{y}$ que está por generar.

ii Use la aplicación para generar 3000 muestras de tamaño 12 y obtener el histograma asociado con las medias muestrales respectivas, $\overline{y}_i$, $i = 1, 2, \ldots, 3000$. ¿Cuál es la forma general de la distribución muestral simulada que obtuvo? Compare la forma de esta distribución muestral simulada con la que obtuvo en el inciso d.

iii Haga clic en el botón "Show Stats" para observar la media y la desviación estándar de los 3000 valores $\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{3000}$. ¿Cómo se comparan estos valores con los que usted pronosticó en el inciso i?

7.5 **Ejercicio Applet** ¿Qué aspecto tiene la distribución de muestreo de la media muestral si las muestras se toman de una distribución aproximadamente normal? Use el applet *Sampling Distribution of the Mean* (en www.thomsonedu.com/statistics/wackerly) para completar lo siguiente. La población de la que se obtendrán las muestras está distribuida aproximadamente en forma normal con $\mu = 16.50$ y $\sigma = 6.03$ (estos valores se proporcionan arriba del histograma poblacional y están denotados como *M* y *S*, respectivamente).

a Use el botón "Next Obs" para seleccionar un solo valor de la población aproximadamente normal. Haga clic cuatro veces en el botón para completar una muestra de tamaño 5. ¿Qué valor obtuvo para la media de esta muestra? Localice este valor en el histograma del fondo (el histograma para los valores de $\overline{Y}$).

b Haga clic en el botón "Reset" para borrar la gráfica del centro. Haga clic en el botón "Next Obs" cinco veces más para obtener otra muestra de tamaño 5 de la población. ¿Qué valor obtuvo para la media de esta nueva muestra? El valor que obtuvo ¿es igual al obtenido en el inciso a? ¿Por qué sí o por qué no?

c Use el botón "1 Sample" ocho veces más para obtener un total de diez valores de la media muestral. Vea el histograma de estas diez medias.

i ¿Qué observa?

ii ¿Cómo se compara la media de estos 10 valores $\overline{y}$ con la media poblacional μ?

d Use el botón "1 Sample" hasta que haya obtenido y graficado 25 valores para la media muestral $\overline{Y}$, cada uno basado en una muestra de tamaño 5.

i ¿Qué observa acerca de la forma del histograma de los 25 valores de $\overline{y}$, $i = 1, 2, \ldots, 25$?

ii ¿Cómo se compara el valor de la desviación estándar de los 25 valores $\overline{y}$ con el valor teórico para $\sigma_{\overline{Y}}$ obtenido en el Ejemplo 5.27, donde demostramos que si $\overline{Y}$ se calcula con base en una muestra de tamaño *n*, entonces $V(\overline{Y}) = \sigma^2/n$?

e Haga clic en el botón "1000 Samples" unas cuantas veces, observando cambios en el histograma a medida que genere más y más valores de la media muestral. ¿Qué observa acerca de la forma del histograma resultante para la distribución muestral simulada de $\overline{Y}$?

f Haga clic en el botón "Toggle Normal" para recubrir (en verde) la distribución normal con la misma media y desviación estándar que el conjunto de valores de $\overline{Y}$ que previamente generó. ¿Esta distribución normal parece ser una buena aproximación de la distribución muestral de $\overline{Y}$?

7.6 **Ejercicio Applet** ¿Cuál es el efecto del tamaño muestral en la distribución muestral de $\overline{Y}$? Use el applet *SampleSize* para completar lo siguiente. Como en el Ejercicio 7.5, la población de la que se obtendrán las muestras está distribuida normalmente en forma aproximada con $\mu = 16.50$ y $\sigma = 6.03$ (estos valores se proporcionan arriba del histograma de población y se denotan como M y S, respectivamente).

a Use las flechas arriba/abajo en la caja izquierda "Sample Size" para seleccionar uno de los tamaños muestrales pequeños disponibles y las flechas de la caja derecha "Sample Size" para seleccionar un tamaño muestral más grande.

b Haga clic en el botón "1 Sample" unas cuantas veces. ¿Qué semejanzas existen entre los dos histogramas que generó? ¿Qué diferencias hay entre ellos?

c Haga clic en el botón "1000 Samples" unas cuantas veces y conteste las preguntas del inciso b.

d Las medias y las desviaciones estándar de las dos distribuciones muestrales ¿están cercanas a los valores que esperaba? [*Sugerencia:* $V(\overline{Y}) = \sigma^2/n$.]

e Haga clic en el botón "Toggle Normal". ¿Qué observa acerca de lo adecuado de la aproximación de las distribuciones normales?

7.7 **Ejercicio Applet** ¿Qué aspecto tiene la distribución de muestreo de la varianza muestral si obtenemos muestras de una población con una distribución aproximadamente normal? Averígüelo usando el applet *Sampling Distribution of the Variance (Mound Shaped Population)* (en www.thomsonedu.com/statistics/wackerly) para completar lo siguiente.

a Haga clic en el botón "Next Obs" para tomar una muestra de tamaño 1 de la población con distribución representada por el histograma de la parte superior. El valor obtenido se grafica en el histograma central. Haga clic cuatro veces más para completar una muestra de tamaño 5. El valor de la varianza muestral se calcula y se proporciona arriba del histograma central. ¿El valor de la varianza muestral es igual al valor de la varianza poblacional? ¿Le sorprende esto?

b Cuando complete el inciso a, el valor de la varianza muestral también se grafica en el histograma de la parte más baja. Haga clic en el botón "Reset" y repita el proceso del inciso a para generar un segundo valor observado para la varianza muestral. ¿Obtuvo el mismo valor que observó en el inciso a? ¿Por qué sí o por qué no?

c Haga clic en el botón "1 Sample" unas cuantas veces. Observará que diferentes muestras llevan a valores diferentes de la varianza muestral. Haga clic en el botón "1000 Samples" unas cuantas veces para generar rápidamente un histograma de los valores observados de la varianza muestral (con base en muestras de tamaño 5). ¿Cuál es la media de los valores de la varianza muestral que generó? ¿La media es cercana al valor de la varianza poblacional?

d En los ejercicios previos de esta sección usted obtuvo distribuciones muestrales simuladas para la media muestral. Todas estas distribuciones muestrales fueron bien aproximadas (para tamaños muestrales grandes) por una distribución normal. Aun cuando la distribución que obtuvo tiene forma de campana, ¿la distribución muestral de la varianza muestral parece ser simétrica (como la distribución normal)?

e Haga clic en el botón "Toggle Theory" para recubrir la función de densidad teórica para la distribución muestral de la varianza de una muestra de tamaño 5 tomada de una población normalmente distribuida. ¿La densidad teórica da una aproximación razonable a los valores representados en el histograma?

f El teorema 7.3, en la sección siguiente, indica que si una muestra aleatoria de tamaño n se toma de una población normalmente distribuida, entonces $(n - 1)S^2/\sigma^2$ tiene una distribución χ^2 con $(n - 1)$ grados de libertad. ¿Este resultado parece consistente con lo que observó en los incisos d y e?

7.8 **Ejercicio Applet** ¿Cuál es el efecto del tamaño de la muestra en la distribución muestral de S^2? Use la aplicación *VarianceSize* para completar lo siguiente. Al igual que en algunos ejercicios previos, la población por muestrear está distribuida normalmente en forma aproximada con $\mu = 16.50$ y $\sigma = 6.03$.

a ¿Cuál es el valor de la varianza poblacional σ^2?

b Use las flechas arriba/debajo de la caja izquierda "Sample Size" para seleccionar uno de los pequeños tamaños muestrales disponibles, y las flechas de la caja derecha "Sample Size" para seleccionar un tamaño muestral más grande.

i Haga clic en el botón "1 Sample" unas pocas veces. ¿Qué hay de semejante en los dos histogramas que generó? ¿Qué es diferente en ellos?

ii Haga clic en el botón "1000 Samples" unas pocas veces y conteste las preguntas del inciso i.

iii Las medias de las dos distribuciones muestrales ¿son cercanas al valor de la varianza poblacional? ¿Cuál de las dos distribuciones muestrales exhibe menor variabilidad?

iv Haga clic en el botón "Toggle Theory". ¿Qué observa acerca de lo adecuado de las distribuciones teóricas que aproximan?

c Seleccione tamaños muestrales de 10 y 50 para una nueva simulación y haga clic en el botón "1000 Samples" unas pocas veces.

i ¿Cuál de las distribuciones muestrales parece ser más semejante a una distribución normal?

ii Consulte el Ejercicio 7.7(f). En el Ejercicio 7.97 usted demostrará que, para un gran número de grados de libertad, la distribución χ^2 puede ser aproximada por una distribución normal. ¿Parece esto razonable con base en su simulación actual?

7.2 Distribuciones muestrales relacionadas con la distribución normal

Ya señalamos que muchos de los fenómenos observados en el mundo real tienen distribuciones de frecuencia relativas que se pueden modelar en forma adecuada con una distribución de probabilidad normal. Por tanto, en muchos problemas prácticos es razonable suponer que las variables aleatorias observables en una muestra aleatoria, $Y_1, Y_2, \ldots, Y_n$, son independientes con la misma función de densidad normal. En el Ejercicio 6.43 se estableció que el estadístico $\overline{Y} = (1/n)(Y_1 + Y_2 + \cdots + Y_n)$ en realidad tiene una distribución normal. Como este resultado se utiliza frecuentemente en nuestras exposiciones subsecuentes, lo presentamos formalmente en el siguiente teorema.

TEOREMA 7.1 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una distribución normal con media μ y varianza σ^2. Entonces

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

está distribuida normalmente con media $\mu_{\overline{Y}} = \mu$ y varianza $\sigma^2_{\overline{Y}} = \sigma^2/n$.

Demostración Como $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2, Y_i, $i = 1, 2, \ldots, n$ son variables independientes distribuidas normalmente, con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$. Además,

$$\begin{aligned}\overline{Y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}(Y_1) + \frac{1}{n}(Y_2) + \cdots + \frac{1}{n}(Y_n)\\ &= a_1Y_1 + a_2Y_2 + \cdots + a_nY_n, \qquad \text{donde } a_i = 1/n,\ i = 1, 2, \ldots, n.\end{aligned}$$

Así, $\overline{Y}$ es una combinación lineal de $Y_1, Y_2, \ldots, Y_n$, y se puede aplicar el Teorema 6.3 para concluir que $\overline{Y}$ está distribuida normalmente con

$$E(\overline{Y}) = E\left[\frac{1}{n}(Y_1) + \cdots + \frac{1}{n}(Y_n)\right] = \frac{1}{n}(\mu) + \cdots + \frac{1}{n}(\mu) = \mu$$

y

$$\begin{aligned}V(\overline{Y}) = V\left[\frac{1}{n}(Y_1) + \cdots + \frac{1}{n}(Y_n)\right] &= \frac{1}{n^2}(\sigma^2) + \cdots + \frac{1}{n^2}(\sigma^2)\\ &= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.\end{aligned}$$

Esto es, la distribución muestral de $\overline{Y}$ es normal con media $\mu_{\overline{Y}} = \mu$ y varianza $\sigma^2_{\overline{Y}} = \sigma^2/n$.

Observe que la varianza de cada una de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ es σ^2 y la varianza de la distribución muestral de la variable aleatoria $\overline{Y}$ es σ^2/n. En lo sucesivo, habrá oportunidad de referirnos a estas dos varianzas. Conservaremos la notación σ^2 para la varianza de las variables aleatorias $Y_1, Y_2, \ldots, Y_n$, y $\sigma^2_{\overline{Y}}$ se usará para denotar la varianza de la distribución muestral de la variable aleatoria $\overline{Y}$. De manera análoga , σ será conservada como la notación para la desviación estándar de las Y_i, y la desviación estándar de la distribución muestral de $\overline{Y}$ se denota $\sigma_{\overline{Y}}$.

De acuerdo con las condiciones del Teorema 7.1, $\overline{Y}$ está normalmente distribuida con media $\mu_{\overline{Y}} = \mu$ y varianza $\sigma^2_{\overline{Y}} = \sigma^2/n$. Se deduce que

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

tiene una distribución normal estándar. Ilustraremos la aplicación del Teorema 7.1 con el siguiente ejemplo.

EJEMPLO **7.2** Una máquina embotelladora puede ser regulada para que descargue un promedio de μ onzas por botella. Se ha observado que la cantidad de líquido dosificado por la máquina está distribuida normalmente con $\sigma = 1.0$ onza. Una muestra de $n = 9$ botellas se selecciona aleatoriamente de la producción de la máquina en un día determinado (todas embotelladas con el mismo ajuste de la máquina) y las onzas de contenido líquido se miden para cada una. Determine la probabilidad de que la media muestral se encuentre a no más de .3 onza de la verdadera media μ para el ajuste seleccionado de la máquina.

Solución Si $Y_1, Y_2, \ldots, Y_9$ denota el contenido en onzas de las botellas que se van a observar, entonces sabemos que las Y_i están distribuidas normalmente con media μ y varianza $\sigma^2 = 1$ para $i = 1, 2, \ldots, 9$. Por tanto, por el Teorema 7.1, $\overline{Y}$ posee una distribución muestral normal con media $\mu_{\overline{Y}} = \mu$ y varianza $\sigma^2_{\overline{Y}} = \sigma^2/n = 1/9$. Deseamos hallar

$$P(|\overline{Y} - \mu| \leq .3) = P[-.3 \leq (\overline{Y} - \mu) \leq .3]$$
$$= P\left(-\frac{.3}{\sigma/\sqrt{n}} \leq \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{.3}{\sigma/\sqrt{n}}\right).$$

Como $(\overline{Y} - \mu_{\overline{Y}})/\sigma_{\overline{Y}} = (\overline{Y} - \mu)/(\sigma/\sqrt{n})$ tiene una distribución normal estándar, se deduce que

$$P(|\overline{Y} - \mu| \leq .3) = P\left(-\frac{.3}{1/\sqrt{9}} \leq Z \leq \frac{.3}{1/\sqrt{9}}\right)$$
$$= P(-.9 \leq Z \leq .9).$$

Usando la Tabla 4, Apéndice 3, encontramos

$$P(-.9 \leq Z \leq .9) = 1 - 2P(Z > .9) = 1 - 2(.1841) = .6318.$$

Por consiguiente, la probabilidad es sólo .6318 de que la media muestral se encuentre a no más de .3 onza de la verdadera media poblacional. ■

EJEMPLO **7.3** Consulte el Ejemplo 7.2. ¿Cuántas observaciones deben estar incluidas en la muestra si deseamos que $\overline{Y}$ se encuentre a no más de .3 onza de μ con probabilidad de .95?

Solución Ahora buscamos

$$P(|\overline{Y} - \mu| \leq .3) = P[-.3 \leq (\overline{Y} - \mu) \leq .3] = .95.$$

Si dividimos cada término de la desigualdad entre $\sigma_{\overline{Y}} = \sigma/\sqrt{n}$ (recuerde que $\sigma = 1$), tenemos

$$P\left[\frac{-.3}{\sigma/\sqrt{n}} \leq \left(\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right) \leq \frac{.3}{\sigma/\sqrt{n}}\right] = P(-.3\sqrt{n} \leq Z \leq .3\sqrt{n}) = .95$$

Pero con el uso de la Tabla 4, Apéndice 3, obtenemos

$$P(-1.96 \leq Z \leq 1.96) = .95.$$

Esto nos dice que

$$.3\sqrt{n} = 1.96 \quad \text{o bien, lo que es equivalente, } n = \left(\frac{1.96}{.3}\right)^2 = 42.68.$$

Desde una perspectiva práctica, es imposible tomar una muestra de tamaño 42.68. Nuestra solución indica que una muestra de tamaño 42 no es suficientemente grande para llegar a nuestro objetivo. Si $n = 43$, $P(|\overline{Y} - \mu| \leq .3)$ es ligeramente mayor que .95. ■

En los capítulos siguientes centraremos nuestra atención en los estadísticos que son funciones de los cuadrados de las observaciones en una muestra aleatoria procedente de una población normal. El Teorema 7.2 establece la distribución muestral de la suma de los cuadrados de variables aleatorias normales estándar e independientes.

TEOREMA 7.2 Si $Y_1, Y_2, \ldots, Y_n$ está definida como en el Teorema 7.1. Entonces $Z_i = (Y_i - \mu)/\sigma$ son variables aleatorias normales estándar e independientes, $i = 1, 2, \ldots, n$, y

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

tienen una distribución χ^2 con n grados de libertad (gl).

Demostración Como $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2, el Ejemplo 6.10 implica que $Z_i = (Y_i - \mu)/\sigma$ tiene una distribución normal estándar para $i = 1, 2, \ldots, n$. Además, las variables aleatorias Z_i son independientes porque las Y_i de las variables aleatorias son independientes, $i = 1, 2, \ldots, n$. El hecho de que $\sum_{i=1}^{n} Z_i^2$ tiene una distribución χ^2 con n grados de libertad se deduce directamente del Teorema 6.4.

En la Tabla 6, Apéndice 3, podemos hallar valores χ^2_α de modo que

$$P\left(\chi^2 > \chi^2_\alpha\right) = \alpha$$

para variables aleatorias con distribuciones χ^2 (véase Figura 7.2). Por ejemplo, si la variable aleatoria de interés χ^2 tiene 10 grados de libertad, la Tabla 6 del Apéndice 3 se puede usar para hallar $\chi^2_{.90}$. Para hacerlo, vea en el renglón marcado 10 gl y la columna con encabezado $\chi^2_{.90}$ y lea el valor 4.86518. Por tanto, si Y tiene una distribución χ^2 con 10 gl, $P(Y > 4.86518) = .90$. Se deduce que $P(Y \le 4.86518) = .10$ y que 4.86518 es el cuantil .10, $\phi_{.10}$, de una variable aleatoria χ^2 con 10 gl. En general,

$$P\left(\chi^2 > \chi^2_\alpha\right) = \alpha \quad \text{implica que} \quad P\left(\chi^2 \le \chi^2_\alpha\right) = 1 - \alpha$$

y que $\chi^2_\alpha = \phi_{1-\alpha}$, el cuantil $(1 - \alpha)$ de la variable aleatoria χ^2.

La Tabla 6, Apéndice 3, contiene $\chi^2_\alpha = \phi_{1-\alpha}$ para diez valores de α (.005, .01, .025, .05, .1, .90, .95, .975, .99 y .995) para cada una de las 37 distribuciones χ^2 diferentes (aquellas con grados de libertad 1, 2, . . . , 30 y 40, 50, 60, 70, 80, 90 y 100). Considerablemente más información acerca de estas distribuciones y la asociada con grados de libertad no incluidos

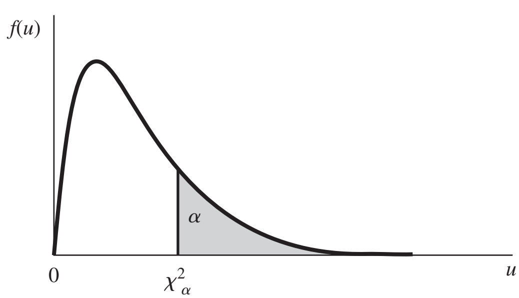


FIGURA **7.2** Una distribución χ^2 que muestra el área α de cola superior en la tabla, se encuentra en software estadístico que se puede adquirir. Si Y tiene una distribución χ^2 con ν grados de libertad, el comando `pchisq(y0,ν)` de R (y S-Plus) da $P(Y \leq y_0)$ mientras que `qchisq(p,ν)` da el p-ésimo *cuantil, el valor* ϕ_p tal que $P(Y \leq \phi_p) = p$. Las probabilidades y los cuantiles asociados con variables aleatorias χ^2 también se pueden obtener fácilmente usando la aplicación *Chi-Square Probabilities and Quantiles* (disponible en www.thomsonedu.com/statistics/wackerly).

El siguiente ejemplo ilustra el uso combinado del Teorema 7.2 y las tablas χ^2.

EJEMPLO **7.4** Si $Z_1, Z_2, \ldots, Z_6$ denota una muestra aleatoria proveniente de la distribución normal estándar, encuentre un número b tal que

$$P\left(\sum_{i=1}^{6} Z_i^2 \leq b\right) = .95.$$

Solución Por el Teorema 7.2, $\sum_{i=1}^{6} Z_i^2$ tiene una distribución χ^2 con 6 grados de libertad. Si vemos la Tabla 6, Apéndice 3, en la fila con encabezado 6 gl y la columna con encabezado $\chi^2_{.05}$, vemos el número 12.5916. Por tanto,

$$P\left(\sum_{i=1}^{6} Z_i^2 > 12.5916\right) = .05, \text{ o bien, lo que es equivalente, } P\left(\sum_{i=1}^{6} Z_i^2 \leq 12.5916\right) = .95,$$

y $b = 12.5916$ es el cuantil .95 (95o. percentil) de la suma de los cuadrados de seis variables aleatorias normales estándar e independientes. ■

La distribución χ^2 desempeña una importante función en muchos procedimientos inferenciales. Por ejemplo, suponga que deseamos hacer una inferencia acerca de la varianza poblacional σ^2 basada en una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de una población normal. Como lo demostraremos en el Capítulo 8, un buen estimador de σ^2 es la varianza muestral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

El siguiente teorema proporciona la distribución de probabilidad para una función del estadístico S^2.

TEOREMA **7.3** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. Entonces

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

tiene una distribución χ^2 con $(n - 1)$ gl. También, $\overline{Y}$ y S^2 son variables aleatorias independientes.

Demostración La demostración completa de este teorema aparece en el Ejercicio 13.93. Para entender mejor el resultado general, consideraremos el caso $n = 2$ y demostraremos que $(n-1)S^2/\sigma^2$ tiene una distribución χ^2 con 1 gl. En el caso de $n = 2$,

$$\overline{Y} = (1/2)(Y_1 + Y_2),$$

y, por tanto,

$$\begin{aligned} S^2 &= \frac{1}{2-1}\sum_{i=1}^{2}(Y_i - \overline{Y})^2 \\ &= \left[Y_1 - \frac{1}{2}(Y_1 + Y_2)\right]^2 + \left[Y_2 - \frac{1}{2}(Y_1 + Y_2)\right]^2 \\ &= \left[\frac{1}{2}(Y_1 - Y_2)\right]^2 + \left[\frac{1}{2}(Y_2 - Y_1)\right]^2 \\ &= 2\left[\frac{1}{2}(Y_1 - Y_2)\right]^2 = \frac{(Y_1 - Y_2)^2}{2}. \end{aligned}$$

Se deduce que, cuando $n = 2$,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(Y_1 - Y_2)^2}{2\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}\right)^2.$$

Demostraremos que esta cantidad es igual al cuadrado de una variable aleatoria normal estándar; es decir, se trata de una variable Z^2 que, como ya hemos demostrado en el Ejemplo 6.11, posee una distribución χ^2 con 1 grado de libertad.

Como $Y_1 - Y_2$ es una combinación de variables aleatorias independientes distribuidas normalmente ($Y_1 - Y_2 = a_1Y_1 + a_2Y_2$ con $a_1 = 1$ y $a_2 = -1$), el Teorema 6.3 nos dice que $Y_1 - Y_2$ tiene una distribución normal con media $1\mu - 1\mu = 0$ y varianza $(1)^2\sigma^2 + (-1)^2\sigma^2 = 2\sigma^2$. Por tanto,

$$Z = \frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}$$

tiene una distribución normal estándar. Como para $n = 2$

$$\frac{(n-1)S^2}{\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2\sigma^2}}\right)^2 = Z^2,$$

se deduce que $(n-1)S^2/\sigma^2$ tiene una distribución χ^2 con 1 grado de libertad.

En el Ejemplo 6.13 demostramos que $U_1 = (Y_1 + Y_2)/\sigma$ y $U_2 = (Y_1 - Y_2)/\sigma$ son variables aleatorias independientes. Observe que, debido a que $n = 2$,

$$\overline{Y} = \frac{Y_1 + Y_2}{2} = \frac{\sigma U_1}{2} \quad \text{y} \quad S^2 = \frac{(Y_1 - Y_2)^2}{2} = \frac{(\sigma U_2)^2}{2}.$$

Como $\overline{Y}$ sólo es una función de U_1 y S^2 es una función de U_2, la independencia de U_1 y U_2 implica la independencia de $\overline{Y}$ y S^2.

EJEMPLO **7.5** En el Ejemplo 7.2, se supone que las onzas de líquido que vierte la máquina embotelladora tienen una distribución normal con $\sigma^2 = 1$. Suponga que planeamos seleccionar una muestra aleatoria de diez botellas y medir la cantidad de líquido en cada una. Si estas diez observaciones se usan para calcular S^2, podría ser útil especificar un intervalo de valores que incluirán S^2 con una probabilidad alta. Encuentre números b_1 y b_2 tales que

$$P(b_1 \leq S^2 \leq b_2) = .90.$$

Solución Observe que

$$P(b_1 \leq S^2 \leq b_2) = P\left[\frac{(n-1)b_1}{\sigma^2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)b_2}{\sigma^2}\right].$$

Debido a que $\sigma^2 = 1$, se deduce que $(n - 1)S^2/\sigma^2 = (n - 1)S^2$ tiene una distribución χ^2 con $(n - 1)$ grados de libertad. Por tanto, podemos usar la Tabla 6, Apéndice 3, para hallar dos números a_1 y a_2 tales que

$$P[a_1 \leq (n-1)S^2 \leq a_2] = .90.$$

Un método para hacer esto es encontrar el valor de a_2 que delimite un área de .05 en la cola superior y el valor de a_1 que delimite .05 en la cola inferior (.95 en la cola superior). Como hay $n - 1 = 9$ grados de libertad, la Tabla 6 del Apéndice 3 indica que $a_2 = 16.919$ y $a_1 = 3.325$. En consecuencia, los valores para b_1 y b_2 que satisfacen nuestras condiciones están dados por

$$3.325 = a_1 = \frac{(n-1)b_1}{\sigma^2} = 9b_1 \quad \text{o} \quad b_1 = \frac{3.325}{9} = .369 \quad \text{y}$$
$$16.919 = a_2 = \frac{(n-1)b_2}{\sigma^2} = 9b_2 \quad \text{o} \quad b_2 = \frac{16.919}{9} = 1.880.$$

Por tanto, si deseamos tener un intervalo que incluya S^2 con probabilidad .90, uno de estos intervalos es (.369, 1.880). Observe que este intervalo es bastante amplio. ■

El resultado del Teorema 7.1 proporciona la base para el desarrollo de procedimientos que permiten hacer inferencias acerca de la media μ de una población normal con varianza conocida σ^2. En dicho caso, el Teorema 7.1 indica que $\sqrt{n}(\overline{Y} - \mu)/\sigma$ tiene una distribución normal estándar. Cuando σ no se conoce, puede ser estimada con $S = \sqrt{S^2}$ y la cantidad

$$\sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right)$$

proporciona la base para desarrollar métodos de inferencia respecto de μ. Demostraremos que $\sqrt{n}(\overline{Y} - \mu)/S$ tiene una distribución conocida como *distribución t de Student* con $n - 1$ grados de libertad. La definición general de una variable aleatoria que posee una distribución t de Student (o simplemente una distribución t) es la siguiente.

DEFINICIÓN **7.2**

> Sea Z una variable aleatoria normal estándar y sea W una variable con distribución χ^2 con ν grados de libertad. Entonces, si W y Z son independientes,
>
> $$T = \frac{Z}{\sqrt{W/\nu}}$$
>
> se dice que tiene una *distribución t* con ν grados de libertad.

Si $Y_1, Y_2, \ldots, Y_n$ constituye una muestra aleatoria de una población normal con media μ y varianza σ^2, el Teorema 7.1 puede aplicarse para demostrar que $Z = \sqrt{n}\,(\overline{Y} - \mu)/\sigma$ tiene una distribución normal estándar. El Teorema 7.3 nos dice que $W = (n - 1)S^2/\sigma^2$ tiene una distribución χ^2 con $\nu = n - 1$ grados de libertad y que Z y W son independientes (puesto que $\overline{Y}$ y S^2 son independientes). Por tanto, según la Definición 7.2,

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\sqrt{n}(\overline{Y} - \mu)/\sigma}{\sqrt{\left[(n-1)S^2/\sigma^2\right]/(n-1)}} = \sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right)$$

tiene una distribución t con $(n - 1)$ grados de libertad.

La ecuación para la función de densidad t no se dará aquí, pero se puede hallar en el Ejercicio 7.98 donde se dan sugerencias acerca de su deducción. Al igual que la función de densidad normal estándar, la función de densidad t es simétrica alrededor de cero. Además, para $\nu > 1$, $E(T) = 0$; y para $\nu > 2$, $V(T) = \nu/(\nu - 2)$. Estos resultados se deducen directamente de los obtenidos en los Ejercicios 4.111 y 4.112 (véase Ejercicio 7.30). De esta manera, vemos que, si $\nu > 1$, una variable aleatoria con distribución t tiene el mismo valor esperado que una variable aleatoria normal estándar. No obstante, una variable aleatoria normal estándar siempre tiene una varianza 1 mientras que, si $\nu > 2$, la varianza de una variable aleatoria con una distribución t siempre es mayor que 1.

La figura 7.3 muestra la gráfica de una función de densidad normal estándar y una función de densidad t. Observe que ambas funciones de densidad son simétricas alrededor del origen pero que la densidad t tiene más masa de probabilidad en sus extremos.

Los valores de t_α tales que $P(T > t_\alpha) = \alpha$ se dan en la Tabla 5, Apéndice 3. Por ejemplo, si una variable aleatoria tiene una distribución t con 21 grados de libertad, $t_{.100}$ se encuentra viendo el renglón marcado 21 gl (grados de libertad) y la columna con encabezado $t_{.100}$. Con el uso de la Tabla 5, vemos que $t_{.100} = 1.323$ y que para 21 grados de libertad, $P(T > 1.323) = .100$. Se deduce que 1.323 es el cuantil .90 (el 90o. percentil) de la distribución t con 21 grados de libertad y en general que $t_\alpha = \phi_{1-\alpha}$, el cuantil $(1 - \alpha)$ [el percentil $100(1 - \alpha)$–ésimo] de una variable aleatoria de distribución t.

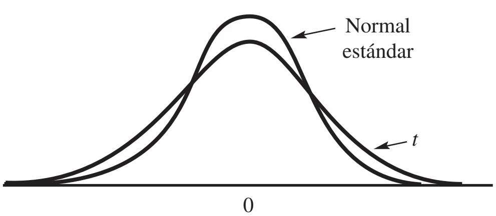


FIGURA **7.3** Comparación de las funciones de densidad normal estándar y t

La Tabla 5, Apéndice 3, contiene $t_\alpha = \phi_{1-\alpha}$ para cinco valores de α (.005, .010, .025, .050 y .100) y 30 distribuciones t diferentes (aquellas con grados de libertad 1, 2, …, 29 e ∞). De manera considerable más información acerca de estas distribuciones y las asociadas con grados de libertad no incluidos en la tabla, es proporcionada por software estadístico comercialmente disponible. Si Y tiene una distribución t con ν grados de libertad, el comando `pt(y0,ν)` de R (y S-Plus) da $P(Y \le y_0)$ mientras que `qt(p,ν)` da el p-ésimo cuantil, el valor de ϕ_p tal que $P(Y \le \phi_p) = p$. Las probabilidades y cuantiles asociados con variables aleatorias con distribución t también se pueden obtener fácilmente utilizando la aplicación breve *Student's Probabilities and Quantiles* (en www.thomsonedu.com/statistics/wackerly).

EJEMPLO 7.6 La resistencia a la tensión para un tipo de alambre está distribuida normalmente con media desconocida μ y varianza desconocida σ^2. Seis trozos de alambre se seleccionan aleatoriamente de un rollo largo; Y_1, la resistencia a la tensión para el trozo i, se mide para $i = 1, 2, \ldots, 6$. La media poblacional μ y la varianza σ^2 pueden ser estimadas por $\overline{Y}$ y S^2, respectivamente. Como $\sigma^2_{\overline{Y}} = \sigma^2/n$, se deduce que $\sigma^2_{\overline{Y}}$ puede ser estimada por S^2/n. Encuentre la probabilidad aproximada de que $\overline{Y}$ esté dentro de $2S/\sqrt{n}$ de la verdadera media poblacional μ.

Solución Deseamos hallar

$$P\left[-\frac{2S}{\sqrt{n}} \le (\overline{Y} - \mu) \le \frac{2S}{\sqrt{n}}\right] = P\left[-2 \le \sqrt{n}\left(\frac{\overline{Y} - \mu}{S}\right) \le 2\right]$$
$$= P(-2 \le T \le 2),$$

donde T tiene una distribución t con, en este caso, $n - 1 = 5$ grados de libertad. Al observar la Tabla 5, Apéndice 3, vemos que el área de la cola superior a la derecha de 2.015 es .05. En consecuencia,

$$P(-2.015 \le T \le 2.015) = .90,$$

y la probabilidad de que $\overline{Y}$ esté a no más de 2 desviaciones estándar estimadas de μ es ligeramente menor que .90. En el Ejercicio 7.24 el valor exacto para $P(-2 \le T \le 2)$ se hallará usando la aplicación *Student's Probabilities and Quantiles* disponible en www.thomsonedu.com/statistics/wackerly.

Observe que si σ^2 se conociera, la probabilidad de que $\overline{Y}$ esté a no más de $2\sigma_{\overline{Y}}$ de μ estaría dada por

$$P\left[-2\left(\frac{\sigma}{\sqrt{n}}\right) \le (\overline{Y} - \mu) \le 2\left(\frac{\sigma}{\sqrt{n}}\right)\right] = P\left[-2 \le \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right) \le 2\right]$$
$$= P(-2 \le Z \le 2) = .9544. \qquad \blacksquare$$

Suponga que queremos comparar las varianzas de dos poblaciones normales con base en información contenida en muestras aleatorias independientes provenientes de las dos poblaciones. Tomemos muestras de tamaño n_1 y n_2 de las dos poblaciones con varianzas σ_1^2 y σ_2^2, respectivamente. Si calculamos S_1^2 de las observaciones en la muestra 1, entonces S_1^2 calcula σ_1^2. Del mismo modo, S_2^2 calculada de las observaciones en la segunda muestra calcula σ_2^2. Entonces, parece que la razón S_1^2/S_2^2 podría usarse para hacer inferencias acerca de las magnitudes relativas de σ_1^2 y σ_2^2. Si dividimos cada S_i^2 entre σ_i^2, entonces la razón resultante

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\left(\frac{S_1^2}{S_2^2}\right)$$

tiene una *distribución F* con $(n_1 - 1)$ grados de libertad en el numerador y $(n_2 - 1)$ grados de libertad en el denominador. La definición general de una variable aleatoria que posee una distribución F aparece a continuación.

DEFINICIÓN **7.3**

Sean W_1 y W_2 variables aleatorias *independientes* con distribución χ^2, con ν_1 y ν_2 grados de libertad, respectivamente. Entonces se dice que

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

tiene una distribución F con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador.

La función de densidad para una variable aleatoria con distribución F se proporciona en el Ejercicio 7.99 donde el método para su deducción está indicado. Se puede demostrar (véase el Ejercicio 7.34) que si F posee una distribución F con ν_1 grados de libertad en el numerador y ν_2 en el denominador, entonces $E(F) = \nu_2/(\nu_2 - 2)$ si $\nu_2 > 2$. También, si $\nu_2 > 4$, entonces $V(F) = [2\nu_2^2(\nu_1 + \nu_2 - 2)]/[\nu_1(\nu_2 - 2)^2(\nu_2 - 4)]$. Observe que la media de una variable aleatoria con distribución F depende sólo del número de grados de libertad ν_2 del denominador.

Considerando una vez más dos muestras aleatorias independientes tomadas de distribuciones normales, sabemos que $W_1 = (n_1 - 1)S_1^2/\sigma_1^2$ y $W_2 = (n_2 - 1)S_2^2/\sigma_2^2$ tienen distribuciones χ^2 independientes con $\nu_1 = (n_1 - 1)$ y $\nu_2 = (n_2 - 1)$ grados de libertad, respectivamente. Entonces, la Definición 7.3 implica que

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} = \frac{\left[(n_1 - 1)S_1^2/\sigma_1^2\right]/(n_1 - 1)}{\left[(n_2 - 1)S_2^2/\sigma_2^2\right]/(n_2 - 1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

tiene una distribución F con $(n_1 - 1)$ grados de libertad en el numerador y $(n_2 - 1)$ grados de libertad en el denominador.

En la Figura 7.4 se muestra la gráfica de una función de densidad F. Los valores de F_α tales que $P(F > F_\alpha) = \alpha$ se dan en la Tabla 7, Apéndice 3, para valores de $\alpha = .100, .050, .025, .010$ y $.005$. En la Tabla 7, los encabezados de las columnas son los grados de libertad del numerador mientras que los grados de libertad del denominador se dan en los encabezados del renglón principal. Opuestos a cada uno de los grados de libertad del denominador (encabezados de renglón), aparecen los valores de $\alpha = .100, .050, .025, .010$ y $.005$. Por ejemplo, si la variable F de interés tiene 5 grados de libertad en el numerador y 7 grados de libertad en el denominador, entonces $F_{.100} = 2.88$, $F_{.050} = 3.97$, $F_{.025} = 5.29$, $F_{.010} = 7.46$ y $F_{.005} = 9.52$. Por tanto, si F tiene una distribución F con 5 grados de libertad en el numerador y 7 grados de

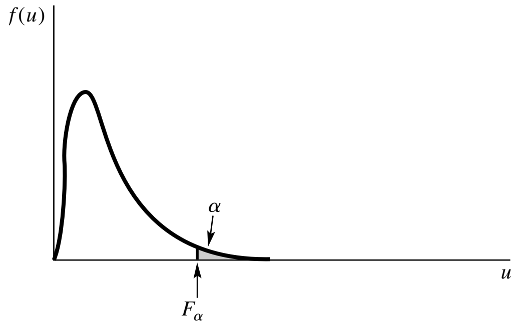


FIGURA **7.4** Una típica función de densidad de probabilidad *F*

libertad en el denominador, entonces $P(F > 7.46) = .01$. Se deduce que 7.46 es el .99 cuantil de la distribución F con 5 grados de libertad en el numerador y 7 grados de libertad en el denominador. En general, $F_\alpha = \phi_{1-\alpha}$, el cuantil $(1 - \alpha)$ [el $100(1 - \alpha)$–ésimo percentil] de una variable aleatoria con distribución F.

Para los cinco valores previamente mencionados de α, la Tabla 7, Apéndice 3, proporciona los valores de F_α para 646 distribuciones F (las de grados de libertad 1, 2, . . . , 10, 12, 15, 20, 24, 30, 40, 60, 120 e ∞ en el numerador y las de grados de libertad 1, 2, . . . , 30, 40, 60, 120 e ∞ en el denominador). De forma considerable hay más información acerca de estas distribuciones, y las asociadas con grados de libertad no incluidas en la tabla, en software estadístico que se puede adquirir comercialmente. Si Y tiene una distribución F con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador, el comando `pf(y₀,ν₁,ν₂)` de *R* (y *S*-Plus) da $P(Y \leq y_0)$ mientras que `qf(p,ν₁,ν₂)` da el p-ésimo cuantil, el valor de ϕ_p tal que $P(Y \leq \phi_p) = p$. Las probabilidades y cuantiles asociados con variables aleatorias de distribución F también se pueden obtener fácilmente con el uso de la aplicación breve *F-Ratio Probabilities and Quantiles* (en www.thomsonedu.com/statistics/wackerly).

EJEMPLO **7.7** Si tomamos muestras independientes de tamaños $n_1 = 6$ y $n_2 = 10$ de dos poblaciones normales con la misma varianza poblacional, encuentre el número b tal que

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = .95.$$

Solución Como $n_1 = 6$, $n_2 = 10$ y las varianzas poblacionales son iguales, entonces

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}$$

tiene una distribución F con $\nu_1 = n_1 - 1 = 5$ grados de libertad en el numerador y $\nu_2 = n_2 - 1 = 9$ grados de libertad en el denominador. Asimismo,

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 1 - P\left(\frac{S_1^2}{S_2^2} > b\right).$$

Por tanto, queremos determinar el número b que delimita un área en el extremo superior de .05 bajo la función de densidad F con 5 grados de libertad en el numerador y 9 grados de libertad en el denominador. Si leemos en la columna 5 y renglón 9 de la Tabla 7, Apéndice 3, vemos que el valor apropiado de b es 3.48.

Aun cuando las varianzas poblacionales son iguales, la probabilidad de que la razón entre las varianzas muestrales sea mayor que 3.48 todavía es .05 (suponiendo tamaños muestrales de $n_1 = 6$ y $n_2 = 10$). ■

Esta sección se ha dedicado a desarrollar distribuciones muestrales de diversos estadísticos calculados mediante el uso de las observaciones de una muestra aleatoria tomada de una población normal (o muestras aleatorias independientes extraídas de dos poblaciones normales). En particular, si $Y_1, Y_2, \ldots, Y_n$ representa una muestra aleatoria de una población normal con media μ y varianza σ^2, hemos visto que $\sqrt{n}(\overline{Y} - \mu)/\sigma$ tiene una distribución normal estándar. Asimismo, $(n-1)S^2/\sigma^2$ tiene una distribución χ^2 y $\sqrt{n}(\overline{Y} - \mu)/S$ tiene una distribución t (ambas con $n-1$ grados de libertad). Si tenemos dos muestras aleatorias independientes de poblaciones normales con varianzas σ_1^2 y σ_2^2, entonces $F = (S_1^2/\sigma_1^2)/(S_2^2/\sigma_2^2)$ tiene una distribución F. Estas distribuciones muestrales harán posible que podamos evaluar las propiedades de procedimientos inferenciales en capítulos posteriores. En la siguiente sección examinamos las aproximaciones a ciertas distribuciones muestrales que pueden ser muy útiles cuando se desconoce la forma exacta de la distribución muestral o cuando es difícil o tedioso usar la distribución muestral exacta para calcular probabilidades.

Ejercicios

7.9 Consulte el Ejemplo 7.2. La cantidad de líquido dosificado por una máquina embotelladora está distribuida normalmente con $\sigma = 1$ onza. Si $n = 9$ botellas se seleccionan aleatoriamente de la producción de la máquina, encontramos que la probabilidad de que la media muestral se encuentre a no más de .3 onza de la verdadera media es .6318. Suponga que $\overline{Y}$ se ha de calcular usando una muestra de tamaño n.

a Si $n = 16$, ¿cuál es $P(|\overline{Y} - \mu| \leq .3)$?

b Encuentre $P(|\overline{Y} - \mu| \leq .3)$ cuando $\overline{Y}$ se ha de calcular usando muestras de tamaños $n = 25$, $n = 36$, $n = 49$ y $n = 64$.

c ¿Qué patrón observa usted entre los valores para $P(|\overline{Y} - \mu| \leq .3)$ que haya contemplado para diversos valores de n?

d ¿Los resultados obtenidos en el inciso b parecen ser consistentes con el resultado obtenido en el Ejemplo 7.3?

7.10 Consulte el Ejercicio 7.9. Suponga ahora que la cantidad de líquido dosificado por la máquina embotelladora está distribuida normalmente con $\sigma = 2$ onzas.

a Si $n = 9$ botellas se seleccionan aleatoriamente de la producción de la máquina, ¿cuál es $P(|\overline{Y} - \mu| \leq .3)$? Compare esto con la respuesta obtenida en el Ejemplo 7.2.

b Encuentre $P(|\overline{Y} - \mu| \leq .3)$ cuando $\overline{Y}$ se ha de determinar usando muestras de tamaños $n = 25$, $n = 36$, $n = 49$ y $n = 64$.

c ¿Qué patrón observa usted entre los valores para $P(|\overline{Y} - \mu| \leq .3)$ que contempló para los diversos valores de n?

d ¿Cómo se comparan las respectivas probabilidades obtenidas en este problema (donde $\sigma = 2$) con las obtenidas en el Ejercicio 7.9 (donde $\sigma = 1$)?

7.11 Un guardabosque, que estudia los efectos de la fertilización en ciertos bosques de pinos en el sureste, está interesado en estimar el promedio de área de la base de los pinos. Al estudiar áreas basales de pinos similares durante muchos años, descubrió que estas mediciones (en pulgadas cuadradas) están distribuidas normalmente con desviación estándar aproxima de 4 pulgadas cuadradas. Si el guardabosque muestrea $n = 9$ árboles, encuentre la probabilidad de que la media muestral se encuentre a no más de 2 pulgadas cuadradas de la media poblacional.

7.12 Suponga que al guardabosque del Ejercicio 7.11 le gustaría que la media muestral estuviera a no más de 1 pulgada cuadrada de la media poblacional, con probabilidad .90. ¿Cuántos árboles debe medir para asegurar este grado de precisión?

7.13 La Environmental Protection Agency se ocupa del problema de establecer criterios para las cantidades de sustancias químicas tóxicas permitidas en lagos y ríos de agua dulce. Una medida común de toxicidad para cualquier contaminante es la concentración de éste que mataría a la mitad de la especie de prueba en un tiempo determinado (por lo general 96 horas para especies de peces). Esta medida se denomina CL50 (concentración letal que mata 50% de la especie de prueba). En muchos estudios, los valores contenidos en el logaritmo natural de mediciones del CL50 están distribuidos normalmente y, en consecuencia, el análisis está basado en datos del ln(CL50).

Estudios de los efectos del cobre en cierta especie de peces (por ejemplo la especie A) muestran que la varianza de mediciones de ln(CL50) es alrededor de .4 con mediciones de concentración en miligramos por litro. Si han de completarse $n = 10$ estudios sobre el CL50 para cobre, encuentre la probabilidad de que la media muestral de ln(CL50) difiera de la verdadera media poblacional en no más de .5.

7.14 Si en el Ejercicio 7.13 deseamos que la media muestral difiera de la media poblacional en no más de .5 con probabilidad .95, ¿cuántas pruebas deben realizarse?

7.15 Suponga que $X_1, X_2, \ldots, X_m$ y $Y_1, Y_2, \ldots, Y_n$ son muestras aleatorias independientes, con las variables X_i distribuidas normalmente con media μ_1 y varianza σ_1^2 y las variables Y_i distribuidas normalmente con media μ_2 y varianza σ_2^2. La diferencia entre las medias muestrales, $\overline{X} - \overline{Y}$, es entonces una combinación lineal de $m + n$ variables aleatorias distribuidas normalmente y, por el Teorema 6.3, tiene una distribución normal.

a Encuentre $E(\overline{X} - \overline{Y})$.
b Encuentre $V(\overline{X} - \overline{Y})$.
c Suponga que $\sigma_1^2 = 2$, $\sigma_2^2 = 2.5$ y $m = n$. Encuentre los tamaños muestrales para que $(\overline{X} - \overline{Y})$ se encuentre a no más de 1 unidad de $(\mu_1 - \mu_2)$ con probabilidad .95.

7.16 Refiriéndose al Ejercicio 7.13, suponga que los efectos del cobre en una segunda especie (por ejemplo la especie B) de peces muestran la varianza de mediciones de ln(CL50) que son de .8. Si las medias poblacionales del ln(CL50) para las dos especies son iguales, encuentre la probabilidad de que, con muestras aleatorias de diez mediciones de cada especie, la media muestral para la especie A sea mayor a la media muestral para la especie B en al menos 1 unidad.

7.17 **Ejercicio Applet** Consulte el Ejemplo 7.4. Use la aplicación breve *Chi-Square Probabilities and Quantiles* para hallar $P\left(\sum_{i=1}^{6} Z_i^2 \leq 6\right)$. (Recuerde que $\sum_{i=1}^{6} Z_i^2$ tiene una distribución χ^2 con 6 grados de libertad.)

7.18 **Ejercicio Applet** Consulte el Ejemplo 7.5. Si $\sigma^2 = 1$ y $n = 10$, use la aplicación *Chi-Square Probabilities and Quantiles* para hallar $P(S^2 \geq 3)$. Recuerde que, con las condiciones dadas previamente, $9S^2$ tiene una distribución χ^2 con 9 grados de libertad.)

7.19 Los amperímetros producidos por un fabricante se venden con la especificación de que la desviación estándar de las lecturas de la aguja no sea mayor que .2 amperes. Uno de estos amperímetros se utilizó para hacer diez lecturas independientes en un circuito de prueba con corriente constante. Si la varianza muestral de estas diez mediciones es .065 y es razonable suponer que las lecturas están distribuidas normalmente, ¿los resultados sugieren que el amperímetro empleado no satisface las especificaciones del mercado? [*Sugerencia:* encuentre la probabilidad aproximada de que la varianza muestral será mayor que .065 si la verdadera varianza poblacional es .04.]

7.20 **a** Si U tiene una distribución χ^2 con ν grados de libertad, encuentre $E(U)$ y $V(U)$.

b Usando los resultados del Teorema 7.3, encuentre $E(S^2)$ y $V(S^2)$ cuando $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2.

7.21 Consulte el Ejercicio 7.13. Suponga que $n = 20$ observaciones se han de tomar a mediciones ln(CL50) y que $\sigma^2 = 1.4$. Denote con S^2 la varianza muestral de las 20 mediciones.

a Encuentre un número b tal que $P(S^2 \leq b) = .975$.

b Encuentre un número a tal que $P(a \leq S^2) = .975$.

c Si a y b son como en los incisos a y b, ¿cuál es $P(a \leq S^2 \leq b)$?

7.22 **Ejercicio Applet** Como ya indicamos en la Definición 4.10, una variable aleatoria Y tiene una distribución χ^2 con ν grados de libertad si y sólo si Y tiene una distribución gamma con $\alpha = \nu/2$ y $\beta = 2$.

a Use la aplicación *Comparison of Gamma Density Functions* para graficar densidades χ^2 con 10, 40 y 80 grados de libertad.

b ¿Qué observa usted acerca de las formas de estas funciones de densidad? ¿Cuál de ellas es más simétrica?

c En el Ejercicio 7.97 usted demostrará que para valores grandes de ν, una variable aleatoria χ^2 tiene una distribución que puede ser aproximada por una distribución normal con $\mu = \nu$ y $\sigma = \sqrt{2\nu}$. ¿Cómo se comparan la media y desviación estándar de la aproximación a la distribución normal con la media y la desviación estándar de la variable aleatoria χ^2 de Y?

d Consulte las gráficas de las densidades χ^2 que obtuvo en el inciso a. En el inciso c dijimos que, si el número de grados de libertad es grande, la distribución χ^2 se puede aproximar con una distribución normal. ¿Le sorprende esto? ¿Por qué?

7.23 **Ejercicio Applet**

a Use la aplicación *Chi-Square Probabilities and Quantiles* para determinar $P[Y > E(Y)]$ cuando Y tiene distribuciones χ^2 con 10, 40 y 80 grados de libertad.

b ¿Qué observó usted acerca de $P[Y > E(Y)]$ cuando aumenta el número de grados de libertad como en el inciso a?

c ¿Cómo se relaciona lo que usted observó en el inciso b con las formas de las densidades χ^2 que obtuvo en el Ejercicio 7.22?

7.24 **Ejercicio Applet** Consulte el Ejemplo 7.6. Suponga que T tiene una distribución t con 5 grados de libertad.

a Use la aplicación *Student's t Probabilities and Quantiles* para hallar la probabilidad exacta de que T sea mayor que 2.

b Use la aplicación *Student's t Probabilities and Quantiles* para hallar la probabilidad exacta de que T sea menor que –2.

c Use la aplicación *Student's t Probabilities and Quantiles* para hallar la probabilidad exacta de que T esté entre –2 y 2.

d Su respuesta al inciso c es considerablemente menor que $0.9544 = P(-2 \leq Z \leq 2)$. Consulte la figura 7.3 y explique por qué esto es como se esperaba.

7.25 **Ejercicio Applet** Suponga que T es una variable aleatoria con distribución t.

a Si T tiene 5 grados de libertad, use la Tabla 5, Apéndice 3, para hallar $t_{.10}$, el valor tal que $P(T > t_{.10}) = .10$. Encuentre $t_{.10}$ usando la aplicación *Student's t Probabilities and Quantiles*.

b Consulte el inciso a. ¿A qué cuantil corresponde $t_{.10}$? ¿A qué percentil?

c Use el applet *Student's t Probabilities and Quantiles* para hallar el valor de $t_{.10}$ para distribuciones t con 30, 60 y 120 grados de libertad.

d Cuando Z tiene una distribución normal estándar, $P(Z > 1.282) = .10$ y $z_{.10} = 1.282$. ¿Qué propiedad de la distribución t (cuando se compara con la distribución normal estándar) explica el hecho de que todos los valores obtenidos en el inciso c son mayores que $z_{.10} = 1.282$?

e ¿Qué se observa acerca de los tamaños relativos de los valores de $t_{.10}$ para distribuciones t con 30, 60 y 120 grados de libertad? Calcule lo que $t_{.10}$ "converge" cuando se hace grande el número de grados de libertad. [*Sugerencia:* vea el renglón marcado ∞ en la Tabla 5, Apéndice 3.]

7.26 Consulte el Ejercicio 7.11. Suponga que, en el problema de fertilización del bosque, la desviación estándar poblacional de áreas basales no se conoce y debe estimarse a partir de la muestra. Si se ha de medir una muestra aleatoria de $n = 9$ áreas basales, encuentre dos estadísticos g_1 y g_2 tales que $P[g_1 \leq (\overline{Y} - \mu) \leq g_2] = .90$.

7.27 **Ejercicio Applet** Consulte el Ejemplo 7.7. Si tomamos muestras independientes de tamaños $n_1 = 6$ y $n_2 = 10$ de dos poblaciones normales con varianzas poblacionales iguales, use la aplicación *F-Ratio Probabilities and Quantiles* para hallar

a $P(S_1^2/S_2^2 > 2)$.

b $P(S_1^2/S_2^2 < 0.5)$.

c la probabilidad de que una de las varianzas muestrales sea al menos el doble de grande que la otra.

7.28 **Ejercicio Applet** Suponga que Y tiene una distribución F con $\nu_1 = 4$ grados de libertad en el numerador y $\nu_2 = 6$ grados de libertad en el denominador.

a Use la Tabla 7, Apéndice 3, para hallar $F_{.025}$. También determínelo con la aplicación *F-Ratio Probabilities and Quantiles.*

b Consulte el inciso a. ¿A qué cuantil de Y corresponde $F_{.025}$? ¿A qué percentil?

c Consulte los incisos a y b. Use la aplicación *F-Ratio Probabilities and Quantiles* para hallar $F_{.975}$, el cuantil .025 (2.5o. percentil) de la distribución de Y.

d Si U tiene una distribución F con $\nu_1 = 6$ grados de libertad en el numerador y $\nu_2 = 4$ en el denominador, use la Tabla 7, Apéndice 3 o la aplicación *F-Ratio Probabilities and Quantiles* para hallar $F_{.025}$.

e En el Ejercicio 7.29, usted demostrará que si Y es una variable aleatoria que tiene una distribución F con ν_1 grados de libertad en el numerador y ν_2 en el denominador, entonces $U = 1/Y$ tiene una distribución F con ν_2 grados de libertad en el numerador y ν_1 en el denominador. ¿Este resultado explica la relación entre $F_{.975}$ del inciso c (4 grados de libertad en el numerador y 6 en el denominador) y $F_{.025}$ del inciso d (6 grados del libertad en el numerador y 4 en el denominador)? ¿Cuál es esta relación?

7.29 Si Y es una variable aleatoria que tiene una distribución F con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador, demuestre que $U = 1/Y$ tiene una distribución F con ν_2 grados de libertad en el numerador y ν_1 grados de libertad en el denominador.

***7.30** Suponga que Z tiene una distribución normal estándar y que Y es una variable aleatoria independiente con distribución χ^2 y con ν grados de libertad. Entonces, según la Definición 7.2,

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

tiene una distribución t con ν grados de libertad.[1]

a Si Z tiene una distribución normal estándar, dé $E(Z)$ y $E(Z^2)$. [*Sugerencia:* para cualquier variable aleatoria, $E(Z^2) = V(Z) + (E(Z))^2$.]

1. Los ejercicios precedidos por un asterisco son opcionales.

b De acuerdo con el resultado obtenido en el Ejercicio 4.112(a), si Y tiene una distribución χ^2 con ν grados de libertad, entonces

$$E(Y^a) = \frac{\Gamma([\nu/2] + a)}{\Gamma(\nu/2)} 2^a, \qquad \text{si } \nu > -2a.$$

Use este resultado, el resultado del inciso a y la estructura de T para demostrar lo siguiente. [*Sugerencia:* recuerde la independencia de Z y Y.]

i $E(T) = 0$, si $\nu > 1$.

ii $V(T) = \nu/(\nu - 2)$, si $\nu > 2$.

7.31 **a** Use la Tabla 7, Apéndice 3, para hallar $F_{.01}$ para variables aleatorias con distribución F, todas con 4 grados de libertad en el numerador, pero con grados de libertad en el denominador de 10, 15, 30, 60, 120 e ∞.

b Consulte el inciso a. ¿Qué observa acerca de los valores de $F_{.01}$ conforme aumenta el número de grados de libertad del denominador?

c ¿Cuál es $\chi^2_{.01}$ para una variable aleatoria con distribución χ^2 y 4 grados de libertad?

d Divida el valor de $\chi^2_{.01}$ (4 grados de libertad) del inciso c entre el valor de $F_{.01}$ (4 grados de libertad en el numerador; grados de libertad $= \infty$ en el denominador). Explique por qué el valor que obtuvo es adecuado para la razón. [*Sugerencia:* considere la definición de una variable aleatoria con distribución F dada en la Definición 7.3.]

7.32 **Ejercicio Applet**

a Encuentre $t_{.05}$ para una variable aleatoria con distribución t y 5 grados de libertad.

b Consulte el inciso a. ¿Cuál es $P(T^2 > t^2_{.05})$?

c Encuentre $F_{.10}$ para una variable aleatoria con distribución F con 1 grado de libertad en el numerador y 5 grados de libertad en el denominador.

d Compare el valor de $F_{.10}$ hallado en el inciso c con el valor de $t^2_{.05}$ de los incisos a y b.

e En el Ejercicio 7.33 usted demostrará que si T tiene una distribución t con ν grados de libertad, entonces $U = T^2$ tiene una distribución F con 1 grado de libertad en el numerador y ν grados del libertad del denominador. ¿Cómo explica esto la relación entre los valores de $F_{.10}$ (1 grado de libertad 1 en el numerador, 5 grados de libertad en el denominador) y $t^2_{.05}$ (5 grados de libertad) que observó en el inciso d?

7.33 Use las estructuras de T y F dadas en las Definiciones 7.2 y 7.3, respectivamente, para demostrar que si T tiene una distribución t con ν grados de libertad, entonces $U = T^2$ tiene una distribución F con 1 grado de libertad en el numerador y ν grados de libertad en el denominador.

7.34 Suponga que W_1 y W_2 son variables aleatorias independientes y con distribución χ^2 con ν_1 y ν_2 grados de libertad, respectivamente. De acuerdo con la Definición 7.3,

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

tiene una distribución F con ν_1 y ν_2 grados de libertad en numerador y denominador, respectivamente. Use la estructura anterior de F, la independencia de W_1 y W_2 y el resultado resumido en el Ejercicio 7.30(b) para demostrar

a $E(F) = \nu_2/(\nu_2 - 2)$, si $\nu_2 > 2$.

b $V(F) = [2\nu_2^2(\nu_1 + \nu_2 - 2)]/[\nu_1(\nu_2 - 2)^2(\nu_2 - 4)]$, si $\nu_2 > 4$.

7.35 Consulte el Ejercicio 7.34. Suponga que F tiene una distribución F con $\nu_1 = 50$ grados de libertad en el numerador y $\nu_2 = 70$ grados de libertad en el denominador. Observe que la Tabla 7, Apéndice 3, no contiene entradas para 50 grados de libertad en el numerador y 70 grados de libertad en el denominador.

a ¿Cuál es $E(F)$?

b Obtenga $V(F)$.

c ¿Es probable que F sea mayor que 3? [*Sugerencia:* use el teorema de Tchebysheff.]

***7.36** Sea S_1^2 la varianza muestral para una muestra aleatoria de diez valores ln(CL50) para cobre y sea S_2^2 la varianza muestral para una muestra aleatoria de ocho valores ln(CL50) para plomo; se utilizaron muestras de la misma especie de peces. Se supone que la varianza poblacional para mediciones de cobre es el doble de la correspondiente varianza poblacional para mediciones de plomo. Suponga que S_1^2 es independiente de S_2^2.

a Encuentre un número b tal que

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = .95.$$

b Encuentre un número a tal que

$$P\left(a \leq \frac{S_1^2}{S_2^2}\right) = .95.$$

[*Sugerencia:* use el resultado del Ejercicio 7.29 y observe que $P(U_1/U_2 \leq k) = P(U_2/U_1 \geq 1/k)$.]

c Si a y b son los de los incisos a y b, encuentre

$$P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right).$$

7.37 Sea $Y_1, Y_2, \ldots, Y_5$ una muestra aleatoria de tamaño 5 de una población normal con media 0 y varianza 1 y sea $\overline{Y} = (1/5)\sum_{i=1}^{5} Y_i$. Sea Y_6 otra observación independiente de la misma población. ¿Cuál es la distribución de

a $W = \sum_{i=1}^{5} Y_i^2$? ¿Por qué?

b $U = \sum_{i=1}^{5} (Y_i - \overline{Y})^2$? ¿Por qué?

c $\sum_{i=1}^{5} (Y_i - \overline{Y})^2 + Y_6^2$? ¿Por qué?

7.38 Suponga que $Y_1, Y_2, \ldots, Y_5, Y_6, \overline{Y}, W$ y U son como se define en el Ejercicio 7.37. ¿Cuál es la distribución de

a. $\sqrt{5}Y_6/\sqrt{W}$? ¿Por qué?

b $2Y_6/\sqrt{U}$? ¿Por qué?

c $2\left(5\overline{Y}^2 + Y_6^2\right)/U$? ¿Por qué?

7.39 Suponga que muestras independientes (de tamaño n_i) se toman de cada una de k poblaciones y que la población i está normalmente distribuida con media μ_i y varianza σ^2, $i = 1, 2, \ldots, k$. Esto es, todas las poblaciones están distribuidas normalmente con la *misma* varianza pero con (posiblemente) medias diferentes. Sean $\overline{X}_i$ y S_i^2, $i = 1, 2, \ldots, k$ las respectivas medias muestrales y varianzas. Sea $\theta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$, donde $c_1, c_2, \ldots, c_k$ son constantes dadas.

a Calcule la distribución de $\hat{\theta} = c_1\overline{X}_1 + c_2\overline{X}_2 + \cdots + c_k\overline{X}_k$. Proporcione razones para cualesquiera afirmaciones que haga.

b Proporcione la distribución de

$$\frac{\text{SSE}}{\sigma^2}, \quad \text{donde SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2.$$

Justifique las afirmaciones que haga.

c Proporcione la distribución de

$$\frac{\hat{\theta}-\theta}{\sqrt{\left(\frac{c_1^2}{n_1}+\frac{c_2^2}{n_2}+\cdots+\frac{c_k^2}{n_k}\right)\text{MSE}}}, \qquad \text{donde MSE} = \frac{\text{SSE}}{n_1+n_2+\cdots+n_k-k}.$$

Justifique las afirmaciones que haga.

7.3 Teorema del límite central

En el Capítulo 5 demostramos que si $Y_1, Y_2, \ldots, Y_n$ representa una muestra aleatoria proveniente de *cualquier* distribución con media μ y varianza σ^2, entonces $E(Y) = \mu$ y $V(\overline{Y}) = \sigma^2/n$. En esta sección desarrollaremos una aproximación para la distribución muestral de $\overline{Y}$, que se puede usar sin considerar la distribución de la población de la cual se tome la muestra.

Si extraemos la muestra de una población normal, el Teorema 7.1 nos dice que $\overline{Y}$ tiene una distribución de muestreo normal. Pero, ¿qué podemos decir acerca de la distribución muestral de $\overline{Y}$ si las variables Y_i no están distribuidas normalmente? Por fortuna, $\overline{Y}$ tendrá una distribución muestral que es aproximadamente normal si el tamaño de la muestra es grande. El enunciado formal de este resultado recibe el nombre de *teorema del límite central*. No obstante, antes de enunciar este teorema, veremos algunos estudios prácticos que demuestran la distribución muestral de $\overline{Y}$.

Se utilizó una computadora para generar muestras aleatorias de tamaño n de una función con densidad exponencial y media 10, es decir, de una población con densidad

$$f(y) = \begin{cases} (1/10)e^{-y/10}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En la Figura 7.5 aparece una gráfica de esta función de densidad. La media muestral se calculó para cada muestra y el histograma de frecuencia relativa para los valores de las medias muestrales para 1000 muestras, cada una de tamaño $n = 5$, se presenta en la Figura 7.6. Observe que la Figura 7.6 presenta un histograma que tiene forma aproximada de campana, pero el histograma está ligeramente sesgado.

La Figura 7.7 es una gráfica de un histograma similar de frecuencia relativa de los valores de la media muestral para 1000 muestras, cada una de tamaño $n = 25$. En este caso, la Figura 7.7 muestra un histograma casi simétrico en forma de campana, que se puede calcular en forma bastante cercana con una función de densidad normal.

FIGURA **7.5** Función de densidad exponencial

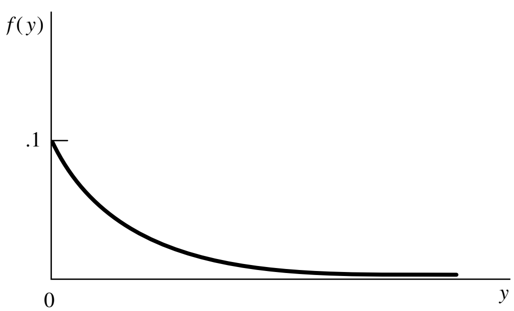


FIGURA 7.6
Histograma de frecuencia relativa: medias muestrales para 1000 muestras (n = 5) tomadas de una distribución exponencial

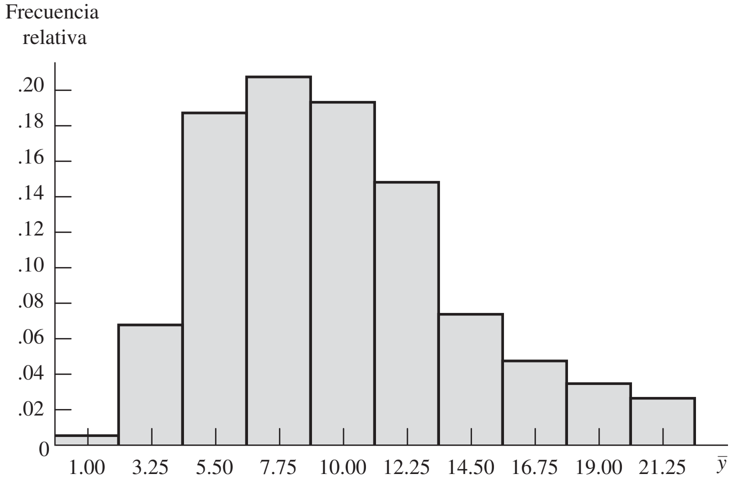
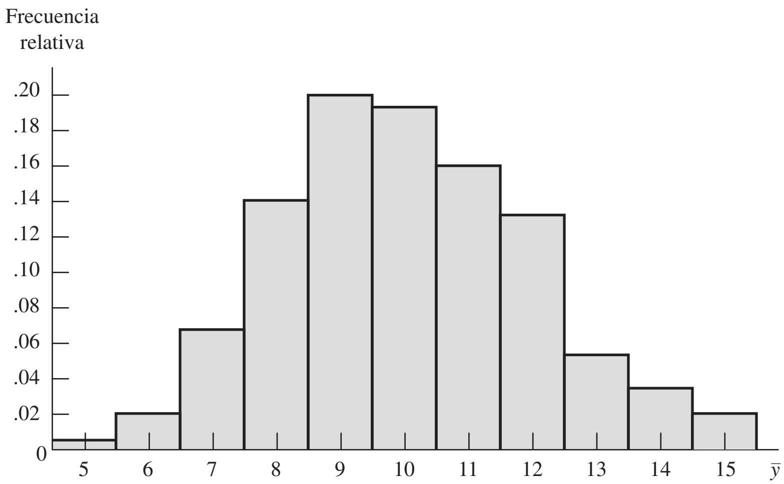


FIGURA 7.7
Histograma de frecuencia relativa: medias muestrales para 1000 muestras (n = 25) tomadas de una distribución exponencial



Recuerde del Capítulo 5 que $E(\overline{Y}) = \mu_{\overline{Y}} = \mu$ y $V(\overline{Y}) = \sigma^2_{\overline{Y}} = \sigma^2/n$. Para la función de densidad exponencial empleada en las simulaciones, $\mu = E(Y_i) = 10$ y $\sigma^2 = V(Y_i) = 10^2 =$ 100. Entonces, para este ejemplo, vemos que

$$\mu_{\overline{Y}} = E(\overline{Y}) = \mu = 10 \quad \text{y} \quad \sigma^2_{\overline{Y}} = V(\overline{Y}) = \frac{\sigma^2}{n} = \frac{100}{n}.$$

Para cada valor de n (5 y 25), calculamos el promedio de las 1000 medias muestrales generadas en el estudio. La varianza observada de las 1000 medias muestrales también se calculó para cada uno de los valores de n. Los resultados se muestran en la Tabla 7.1. En cada estudio práctico ($n = 5$ y $n = 25$), el promedio de las medias muestrales observadas y la varianza de las medias muestrales observadas son bastante cercanos a los valores teóricos.

A continuación enunciamos formalmente el teorema del límite central.

Tabla 7.1 Cálculos para 1000 medias muestrales

Tamaño muestral	Promedio de 1000 medias muestrales	$\mu_{\overline{Y}} = \mu$	Varianza de 1000 medias muestrales	$\sigma^2_{\overline{Y}} = \sigma^2/n$
$n = 5$	9.86	10	19.63	20
$n = 25$	9.95	10	3.93	4

TEOREMA **7.4**

Teorema del límite central: Sean $Y_1, Y_2, \dots, Y_n$ variables aleatorias independientes y distribuidas idénticamente con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Definamos

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \qquad \text{donde } \overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

Entonces la función de distribución de U_n converge hacia la función de distribución normal estándar cuando $n \to \infty$. Esto es,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad \text{para toda } u.$$

El teorema del límite central implica que los enunciados de probabilidad acerca de U_n pueden ser aproximados por probabilidades correspondientes para la variable aleatoria normal estándar si n es grande. (Por lo general, un valor de n mayor que 30 asegura que la distribución de U_n se puede calcular en forma aproximada por medio de una distribución normal.)

Por comodidad, la conclusión del teorema del límite central a menudo se sustituye con el enunciado más sencillo de que $\overline{Y}$ está *distribuida normalmente en forma asintótica* con media μ y varianza σ^2/n. El teorema del límite central se puede aplicar a una muestra aleatoria $Y_1, Y_2, \dots, Y_n$ para *cualquier* distribución mientras $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$ sean finitas y el tamaño muestral sea grande.

Daremos algunos ejemplos de la aplicación del teorema del límite central pero aplazamos la demostración hasta la siguiente sección (cuyo estudio es opcional). La demostración no es necesaria para entender las aplicaciones del teorema del límite central que aparece en este texto.

EJEMPLO **7.8** Las calificaciones de exámenes para todos los estudiantes de último año de preparatoria en cierto estado tienen media de 60 y varianza de 64. Una muestra aleatoria de $n = 100$ estudiantes de una escuela preparatoria grande tuvo una calificación media de 58. ¿Hay evidencia para sugerir que el nivel de conocimientos de esta escuela sea inferior? (Calcule la probabilidad de que la media muestral sea a lo sumo 58 cuando $n = 100$.)

Solución Denote con $\overline{Y}$ la media de una muestra aleatoria de $n = 100$ calificaciones de una población con $\mu = 60$ y $\sigma^2 = 64$. Deseamos calcular $P(\overline{Y} \le 58)$. Sabemos por el Teorema 7.4 que $(\overline{Y} - \mu)/(\sigma/\sqrt{n})$ tiene una distribución que puede aproximarse con una distribución normal estándar. En consecuencia, usando la Tabla 4, Apéndice 3, tenemos

$$P(\overline{Y} \le 58) = P\left(\frac{\overline{Y} - 60}{8/\sqrt{100}} \le \frac{58 - 60}{.8}\right) \approx P(Z \le -2.5) = .0062.$$

Debido a que esta probabilidad es muy pequeña, no es probable que la muestra de la escuela estudiada se pueda considerar como muestra aleatoria de una población con $\mu = 60$ y $\sigma^2 = 64$. La evidencia sugiere que la calificación promedio para esta preparatoria es menor que el promedio general de $\mu = 60$.

Este ejemplo ilustra el uso de probabilidad en el proceso de comprobación de hipótesis, técnica común de inferencia estadística que se estudiará con más detalle en el Capítulo 10. ■

EJEMPLO **7.9** Los tiempos de servicio para los clientes que pasan por la caja en una tienda de venta al menudeo son variables aleatorias independientes con media de 1.5 minutos y varianza de 1.0. Calcule la probabilidad de que 100 clientes puedan ser atendidos en menos de 2 horas de tiempo total de servicio.

Solución Si denotamos con Y_i el tiempo de servicio para el i-ésimo cliente, entonces queremos calcular

$$P\left(\sum_{i=1}^{100} Y_i \leq 120\right) = P\left(\overline{Y} \leq \frac{120}{100}\right) = P(\overline{Y} \leq 1.20).$$

Como el tamaño muestral es grande, el teorema del límite central nos dice que $\overline{Y}$ está distribuida normalmente en forma aproximada con media $\mu_{\overline{Y}} = \mu = 1.5$ y varianza $\sigma^2_{\overline{Y}} = \sigma^2/n = 1.0/100$. Por tanto, usando la Tabla 4, Apéndice 3, tenemos

$$\begin{aligned} P(\overline{Y} \leq 1.20) &= P\left(\frac{\overline{Y} - 1.50}{1/\sqrt{100}} \leq \frac{1.20 - 1.50}{1/\sqrt{100}}\right) \\ &\approx P[Z \leq (1.2 - 1.5)10] = P(Z \leq -3) = .0013. \end{aligned}$$

Entonces, la probabilidad de que 100 clientes puedan ser atendidos en menos de 2 horas es aproximadamente .0013. Esta pequeña probabilidad indica que es prácticamente imposible atender a 100 clientes en menos de 2 horas. ■

Ejercicios

7.40 **Ejercicio Applet** Suponga que la población de interés no tiene una distribución normal. ¿Qué aspecto presenta la distribución muestral de $\overline{Y}$ y cuál es el efecto del tamaño de la muestra en la distribución muestral de $\overline{Y}$? Use la aplicación *SampleSize* para completar lo siguiente. Use las flechas de arriba/abajo a la izquierda del histograma de la distribución poblacional para seleccionar la distribución “Sesgada”. ¿Cuál es la media y la desviación estándar de la población de la cual se seleccionaron las muestras? [Estos valores están marcados M y S, respectivamente, y se dan arriba del histograma poblacional.]

a Use las flechas arriba/abajo en las cajas izquierda y derecha de “Sample Size” para seleccionar muestras de tamaños 1 y 3. Haga clic unas cuantas veces en el botón “1 Sample”. ¿Qué semejanzas hay entre los dos histogramas generados? ¿Qué es diferente en ellos?

b Haga clic unas cuantas veces en el botón “1000 Samples” y conteste las preguntas del inciso b. ¿Los histogramas generados tienen las formas que esperaba? ¿Por qué?

c ¿Las medias y desviaciones estándar de las dos distribuciones muestrales son cercanas a los valores que esperaba? [*Sugerencia:* $V(\overline{Y}) = \sigma^2/n$.]

d Haga clic en el botón “Toggle Normal”. ¿Qué observa acerca de lo adecuado de las distribuciones normales calculadas?

e Haga clic en las dos distribuciones muestrales generadas para que aparezcan ventanas para cada una. Use las flechas arriba/abajo en las cajas izquierda y derecha de “Sample Size” para seleccionar muestras de tamaño 10 y 25. Haga clic en el botón “Toggle Normal”. Ahora tiene ya las gráficas de las distribuciones muestrales de las medias muestrales basadas en muestras de tamaños 1, 3, 10 y 25. ¿Qué observa acerca de lo adecuado de la aproximación normal cuando aumenta el tamaño muestral?

7.41 **Ejercicio Applet** Consulte el Ejercicio 7.40. Use la aplicación *SampleSize* para completar lo siguiente. Use la flecha arriba/abajo a la izquierda del histograma de la distribución poblacional para seleccionar la distribución “en forma de U”. ¿Cuál es la media y la desviación estándar de la población de la cual se seleccionarán las muestras?

a Conteste las preguntas de los incisos (a-e) del Ejercicio 7.40.

b Consulte el inciso a. Cuando examinó la distribución muestral de $\overline{Y}$ para $n = 3$, tenía un “valle” en el centro. ¿Por qué ocurrió esto? Use la aplicación *Basic* para averiguarlo. Seleccione la distribución poblacional “en forma de U” y haga clic en el botón “1 Sample.” ¿Qué observa acerca de los valores de observaciones individuales en la muestra? Haga clic en el botón “1 Sample” varias veces más. ¿Los valores de la muestra tienden a ser (relativamente) grandes o pequeños con pocos valores en el “centro”? ¿Por qué? ¿Qué efecto tiene esto en el valor de la media muestral? [*Sugerencia:* 3 es un tamaño muestral impar.]

7.42 La resistencia a la ruptura del vidrio templado promedia 14 (medida en miles de libras por pulgada cuadrada) y tiene una desviación estándar de 2.

a ¿Cuál es la probabilidad de que el promedio de resistencia a la ruptura de 100 piezas seleccionadas aleatoriamente de este vidrio exceda de 14.5?

b Encuentre un intervalo que incluya, con probabilidad 0.95, el promedio de resistencia a la ruptura de 100 piezas de este vidrio seleccionadas aleatoriamente.

7.43 Una antropóloga desea calcular el promedio de estatura de los hombres de cierta raza. Si se supone que la desviación estándar poblacional es de 2.5 pulgadas y si ella muestrea 100 hombres aleatoriamente, encuentre la probabilidad de que la diferencia entre la media muestral y la verdadera media poblacional no exceda de .5 pulgada.

7.44 Suponga que la antropóloga del Ejercicio 7.43 desea que la diferencia entre la media muestral y la media poblacional sea menor que .4 pulgada, con probabilidad de .95. ¿Cuántos hombres debe tomar como muestra para lograr este objetivo?

7.45 Trabajadores de una gran empresa de servicios tienen un salario promedio de \$7.00 por hora con una desviación estándar de \$.50. La industria tiene 64 trabajadores de cierto grupo étnico que tienen un salario promedio de \$6.90 por hora. ¿Es razonable suponer que la tasa salarial del grupo étnico es equivalente a la de una muestra aleatoria de trabajadores tomada de los empleados en la industria militar? [*Sugerencia:* calcule la probabilidad de obtener una media muestral menor o igual que \$6.90 por hora.]

7.46 La acidez de los suelos se mide mediante una cantidad llamada pH, que varía de 0 (acidez alta) a 14 (alcalinidad alta). Un edafólogo desea calcular el promedio de pH para un campo de grandes dimensiones al seleccionar aleatoriamente n muestras de núcleos y medir el pH de cada muestra. Aun cuando la desviación estándar poblacional de mediciones de pH no se conoce, la experiencia del pasado indica que casi todos los suelos tienen un valor de pH de entre 5 y 8. Si el científico selecciona $n = 40$ muestras, encuentre la probabilidad aproximada de que la media muestral de las 40 mediciones de pH esté a .2 unidades del verdadero promedio de pH para el campo. [*Sugerencia:* vea el Ejercicio 1.17.]

7.47 Suponga que al científico del Ejercicio 7.46 le gustaría que la media muestral estuviera a no más de .1 de la verdadera media con probabilidad .90. ¿Cuántas muestras de núcleos debe tomar?

7.48 Un aspecto importante de un plan económico federal era que los consumidores ahorraran una parte importante de dinero que recibieran por una reducción de impuestos sobre sus ingresos. Suponga que las primeras estimaciones de la parte del total de impuesto ahorrada, con base en una muestra aleatoria de 35 economistas, tuvo media de 26% y desviación estándar de 12%.

a ¿Cuál es la probabilidad aproximada de que la estimación de la media muestral, basada en una muestra aleatoria de $n = 35$ economistas, se encuentre a no más de 1% de la media de la población de las estimaciones de todos los economistas?

b ¿Es necesariamente verdadero que la media de la población de las estimaciones de todos los economistas sea igual al porcentaje de ahorro en impuestos que en realidad se logrará?

7.49 El tiempo necesario para el mantenimiento periódico de un automóvil u otra máquina tiene por lo general una distribución de probabilidad en forma de campana. Debido a que se presentarán algunos alargamientos en los tiempos de servicio, la distribución tiende a estar sesgada a la derecha. Suponga que el tiempo necesario para dar servicio a un automóvil que ha recorrido 5000 millas tiene una media de 1.4 horas y desviación estándar de .7 horas. Suponga también que el departamento de servicio planea atender a 50 automóviles por jornada de 8 horas y que, para hacerlo, puede dedicar un tiempo promedio máximo de sólo 1.6 horas por automóvil. ¿Cuántos días tendrá que trabajar tiempo extra el departamento de servicio?

7.50 Se ha encontrado que las mediciones de resistencia al corte en soldaduras por puntos tienen una desviación estándar de 10 libras por pulgada cuadrada (psi). Si se han de medir 100 soldaduras de prueba, ¿cuál es la probabilidad aproximada de que la media muestral se encuentre a no más de 1 psi de la verdadera media poblacional?

7.51 Consulte el Ejercicio 7.50. Si la desviación estándar de mediciones de resistencia al corte en soldaduras por puntos es 10 psi, ¿cuántas soldaduras de prueba deben muestrearse si deseamos que la media muestral se encuentre a no más de 1 psi de la verdadera media con probabilidad aproximada de .99?

7.52 Los resistores que se han de usar en un circuito tienen un promedio de resistencia de 200 ohms y desviación estándar de 10 ohms. Suponga que 25 de estos resistores se seleccionan aleatoriamente para usarse en un circuito.

a ¿Cuál es la probabilidad de que la resistencia promedio para los 25 resistores esté entre 199 y 202 ohms?

b Encuentre la probabilidad de que la resistencia *total* no exceda de 5100 ohms. [*Sugerencia:* vea el Ejemplo 7.9.]

7.53 Concentraciones de monóxido de carbono de cierta hora en muestras de aire de una gran ciudad promedian 12 ppm (partes por millón) con desviación estándar de 9 partes por millón.

a ¿Cree usted que las concentraciones de monóxido de carbono en las muestras de aire de esta ciudad están distribuidas normalmente? ¿Por qué sí o por qué no?

b Encuentre la probabilidad de que la concentración promedio en 100 muestras seleccionadas aleatoriamente exceda de 14 partes por millón.

7.54 Asfaltos no alterados, como se encuentran por lo general en depósitos de plomo y zinc, tienen razones atómicas de hidrógeno/carbono (H/C) que promedian 1.4 con desviación estándar de .05. Encuentre la probabilidad de que la razón promedio de H/C sea menor que 1.3 si seleccionamos al azar 25 muestras de asfaltos.

7.55 El tiempo de inactividad por día para una central de cómputo tiene una media de 4 horas y desviación estándar de .8 hora.

a Suponga que deseamos calcular probabilidades acerca del promedio diario de inactividad durante un periodo de 30 días.

i ¿Qué suposiciones deben ser verdaderas para usar el resultado del Teorema 7.4 y así obtener una aproximación válida para probabilidades acerca del promedio diario de inactividad?

ii De acuerdo con las suposiciones descritas en el inciso i, ¿cuál es la probabilidad aproximada de que el promedio diario de inactividad, durante un periodo de 30 días, sea de entre 1 y 5 horas?

b De acuerdo con las suposiciones descritas en el inciso a, ¿cuál es la probabilidad aproximada de que el tiempo *total* de inactividad, durante un periodo de 30 días, sea menor que 115 horas?

7.56 Para muchos productos a granel, por ejemplo mineral de hierro y azúcar sin refinar, se muestrea su calidad mediante un método que requiere que pequeñas muestras se tomen periódicamente cuando el material se mueve en una banda transportadora. Las pequeñas muestras se combinan entonces y se mezclan para formar una muestra compuesta. Denote con Y_i el volumen de la i-ésima pequeña muestra de un lote particular y suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria y que cada valor de Y_i tiene una media de μ (en pulgadas cúbicas) y varianza σ^2. El volumen promedio μ de las muestras se puede establecer al ajustar el tamaño del equipo de muestreo. Suponga que se sabe que la varianza σ^2 de los volúmenes de las muestras es aproximadamente 4. El volumen total de la muestra compuesta debe exceder de 200 pulgadas cúbicas con probabilidad aproximadamente de .95 cuando se seleccionan $n = 50$ pequeñas muestras. Determine un ajuste para μ que permita que se satisfagan los requisitos del muestreo.

7.57 Se conectan 25 lámparas de calor en un invernadero para que cuando falle una de ellas, otra tome su lugar de inmediato. (Sólo una lámpara se enciende a la vez.) Las lámparas operan de manera independiente y cada una tiene una vida útil de 50 horas y desviación estándar de 4 horas. Si el invernadero no se revisa durante 1300 horas después de encender el sistema de lámparas, ¿cuál es la probabilidad de que una lámpara permanezca encendida al final del periodo de 1300 horas?

7.58 Suponga que $X_1, X_2, \ldots, X_n$ y que $Y_1, Y_2, \ldots, Y_n$ son muestras aleatorias independientes de poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Demuestre que la variable aleatoria

$$U_n = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

satisface las condiciones del Teorema 7.4 y por tanto que la función de distribución U_n converge hacia una función de distribución normal estándar cuando $n \to \infty$. [*Sugerencia:* considere $W_i = X_i - Y_i$, para $i = 1, 2, \ldots, n$.]

7.59 Se diseña un experimento para determinar si el operador A o el operador B obtienen el trabajo de operar una nueva máquina. Se toma el tiempo a cada operador en 50 intentos independientes que comprenden la realización de cierto trabajo usando la máquina. Si las medias muestrales para los 50 intentos difieren en más de 1 segundo, el operador con el menor tiempo medio obtiene el trabajo. De otro modo, el experimento es considerado como terminado en empate. Si las desviaciones estándar de los tiempos para ambos operadores se suponen de 2 segundos, ¿cuál es la probabilidad de que el operador A obtenga el trabajo aun cuando ambos operadores tengan igual capacidad?

7.60 El resultado del Ejercicio 7.58 se cumple incluso si difieren los tamaños muestrales. Esto es, si $X_1, X_2, \ldots, X_{n_1}$ y $Y_1, Y_2, \ldots, Y_{n_2}$ constituyen muestras aleatorias independientes de poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, entonces $\overline{X} - \overline{Y}$ estará distribuida normalmente en forma aproximada, para n_1 y n_2 grandes, con media $\mu_1 - \mu_2$ y varianza $(\sigma_1^2/n_1) + (\sigma_2^2/n_2)$.

La filtración de agua por el suelo depende, entre otras cosas, de la porosidad (proporción de huecos por volumen) del suelo. Para comparar dos tipos de suelo arenoso, se han de tomar $n_1 = 50$ mediciones de la porosidad del suelo A y $n_2 = 100$ mediciones del suelo B.

Suponga que $\sigma_1^2 = .01$ y $\sigma_2^2 = .02$. Encuentre la probabilidad de que la diferencia entre las medias muestrales esté a no más de .05 unidades de la diferencia entre las medias poblacionales $\mu_1 - \mu_2$.

7.61 Consulte el Ejercicio 7.60. Suponga que $n_1 = n_2 = n$, y encuentre el valor de n que permita que la diferencia entre las medias muestrales sea no mayor que .04 unidades de $\mu_1 - \mu_2$ con probabilidad .90.

7.62 Los tiempos que una cajera emplea para procesar el pedido de un cliente son variables aleatorias independientes, con media de 2.5 minutos y desviación estándar de 2 minutos. ¿Cuál es la probabilidad aproximada de que tome más de 4 horas procesar los pedidos de 100 personas?

7.63 Consulte el Ejercicio 7.62. Encuentre el número de clientes n tal que sea aproximadamente .1 la probabilidad de que los pedidos de los n clientes se puedan procesar en menos de 2 horas.

7.4 Una demostración del teorema del límite central (opcional)

Esbozaremos una demostración del teorema del límite central para el caso en el que existan las funciones generadoras de momento para las variables aleatorias de la muestra. La demostración depende de un resultado fundamental de la teoría de probabilidades que no se puede demostrar aquí pero que se expresa en el Teorema 7.5.

TEOREMA 7.5 Sean Y y $Y_1, Y_2, Y_3 \ldots$ variables aleatorias con funciones generadoras de momento $m(t)$ y $m_1(t), m_2(t), m_3(t), \ldots,$ respectivamente. Si

$$\lim_{n \to \infty} m_n(t) = m(t) \qquad \text{para toda } t \text{ real,}$$

entonces la función de distribución de Y_n converge hacia la función de distribución de Y cuando $n \to \infty$.

A continuación damos la demostración del teorema del límite central, Teorema 7.4.

Demostración Sea

$$\begin{aligned} U_n &= \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i, \qquad \text{donde } Z_i = \frac{Y_i - \mu}{\sigma}. \end{aligned}$$

Debido a que las Y_i de las variables aleatorias son independientes e idénticamente distribuidas, Z_i, $i = 1, 2, \ldots, n$, son independientes e idénticamente distribuidas con $E(Z_i) = 0$ y $V(Z_i) = 1$.

Como la función generadora de momento de la suma de variables aleatorias independientes es el producto de las funciones generadoras de momentos de cada una,

$$m_{\sum Z_i}(t) = m_{Z_1}(t) \times m_{Z_2}(t) \times \cdots \times m_{Z_n}(t) = [m_{Z_1}(t)]^n$$

y

$$m_{U_n}(t) = m_{\sum Z_i}\left(\frac{t}{\sqrt{n}}\right) = \left[m_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Por el teorema de Taylor, con residuo (vea su texto *de Cálculo*)

$$m_{Z_1}(t) = m_{Z_1}(0) + m'_{Z_1}(0)t + m''_{Z_1}(\xi)\frac{t^2}{2}, \qquad \text{donde } 0 < \xi < t,$$

y como $m_{Z_1}(0) = E(e^{0Z_1}) = E(1) = 1$, y $m'_{Z_1}(0) = E(Z_1) = 0$,

$$m_{Z_1}(t) = 1 + \frac{m''_{Z_1}(\xi)}{2}t^2, \qquad \text{donde, } 0 < \xi < t.$$

Por tanto

$$\begin{aligned} m_{U_n}(t) &= \left[1 + \frac{m''_{Z_1}(\xi_n)}{2}\left(\frac{t}{\sqrt{n}}\right)^2\right]^n \\ &= \left[1 + \frac{m''_{Z_1}(\xi_n)t^2/2}{n}\right]^n, \qquad \text{donde } 0 < \xi_n < \frac{t}{\sqrt{n}}. \end{aligned}$$

Observe que conforme $n \to \infty$, $\xi_n \to 0$ y $m''_{Z_1}(\xi_n)t^2/2 \to m''_{Z_1}(0)t^2/2 = E(Z_1^2)t^2/2 = t^2/2$ porque $E(Z_1^2) = V(Z_1) = 1$. Recuerde que si

$$\lim_{n\to\infty} b_n = b \quad \text{entonces.} \quad \lim_{n\to\infty}\left(1 + \frac{b_n}{n}\right)^n = e^b.$$

Finalmente,

$$\lim_{n\to\infty} m_{U_n}(t) = \lim_{n\to\infty}\left[1 + \frac{m''_{Z_1}(\xi_n)t^2/2}{n}\right]^n = e^{t^2/2},$$

es la función generadora de momento para una variable aleatoria normal estándar. Con la aplicación del Teorema 7.5 concluimos que U_n tiene una función de distribución que converge hacia la función de distribución de la variable aleatoria normal estándar.

7.5 Aproximación normal a la distribución binomial

El teorema del límite central también se puede usar para aproximar probabilidades de algunas variables aleatorias discretas cuando las probabilidades exactas sean difíciles de calcular. Un ejemplo útil comprende la distribución binomial para valores grandes del número de intentos n.

Suponga que Y tiene una distribución binomial con n intentos y probabilidad de éxito en cualquier intento denotado por p. Si deseamos hallar $P(Y \leq b)$, podemos usar la función de probabilidad binomial para calcular $P(Y = y)$ para cada entero no negativo y menor o igual a b y luego sumar estas probabilidades. Se puede disponer de tablas para algunos valores del tamaño muestral n, pero el cálculo directo es engorroso para valores grandes de n para los que no hay tablas.

De manera alternativa, podemos ver a Y, el número de éxitos en n intentos, como una suma de una muestra formada de ceros y unos; esto es,

$$Y = \sum_{i=1}^{n} X_i,$$

donde

$$X_i = \begin{cases} 1, & \text{si el } i\text{-ésimo intento resulta en éxito,} \\ 0, & \text{en otro caso.} \end{cases}$$

Las variables aleatorias X_i para $i = 1, 2, \ldots, n$ son independientes (porque los intentos son independientes) y es fácil demostrar que $E(X_i) = p$ y $V(X_i) = p(1 - p)$ para $i = 1, 2, \ldots, n$. En consecuencia, cuando n es grande, la fracción muestral de éxitos,

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

posee una distribución muestral aproximadamente normal con media $E(X_i) = p$ y varianza $V(X_i)/n = p(1 - p)/n$.

Entonces, hemos empleado el Teorema 7.4 (el teorema del límite central) para establecer que si Y es una variable aleatoria binomial con parámetros n y p y si n es grande, entonces Y/n tiene aproximadamente la misma distribución que U, donde U está distribuida normalmente con media $\mu_U = p$ y varianza $\sigma_U^2 = p(1 - p)/n$. De la misma manera, para n grande, podemos considerar que Y tiene aproximadamente la misma distribución que W, donde W está distribuida normalmente con media $\mu_W = np$ y varianza $\sigma_W^2 = np(1 - p)$.

EJEMPLO **7.10** La candidata A piensa que puede ganar las elecciones en una ciudad si obtiene por lo menos 55% de los votos en el distrito electoral 1. También piensa que alrededor de 50% de los votantes de la ciudad están a su favor. Si $n = 100$ votantes se presentan a votar en el distrito electoral 1, ¿cuál es la probabilidad de que la candidata A reciba al menos 55% de sus votos?

Solución Sea Y el número de votantes del distrito electoral 1 que están a favor de la candidata A. Debemos calcular $P(Y/n \geq .55)$ cuando p es la probabilidad de que un votante seleccionado aleatoriamente del distrito electoral 1 esté a favor de la candidata A. Si consideramos los $n =$ 100 votantes del distrito electoral 1 como una muestra aleatoria de la ciudad, entonces Y tiene una distribución binomial con $n = 100$ y $p = .5$. Hemos visto que la fracción de votantes que están a favor de la candidata A es

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

donde $X_i = 1$ si el i-ésimo votante está a favor de la candidata A y $X_i = 0$ de otro modo.

Como es razonable suponer que $X_i = 1, 2, \ldots, n$ son independientes, el teorema del límite central implica que $\overline{X} = Y/n$ está distribuida normalmente en forma aproximada con media

$p = .5$ y varianza $pq/n = (.5)(.5)/100 = .0025$. Por tanto,

$$P\left(\frac{Y}{n} \geq .55\right) = P\left(\frac{Y/n - .5}{\sqrt{.0025}} \geq \frac{.55 - .50}{.05}\right) \approx P(Z \geq 1) = .1587$$

de la Tabla 4, Apéndice 3. ■

La aproximación normal a las probabilidades binomiales funciona bien para las n relativamente grandes mientras p no sea cercana a cero o uno. Una regla práctica útil es que la aproximación normal a la distribución binomial es apropiada cuando $p \pm 3\sqrt{pq/n}$ está en el intervalo (0, 1), es decir, si

$$0 < p - 3\sqrt{pq/n} \quad \text{y} \quad p + 3\sqrt{pq/n} < 1.$$

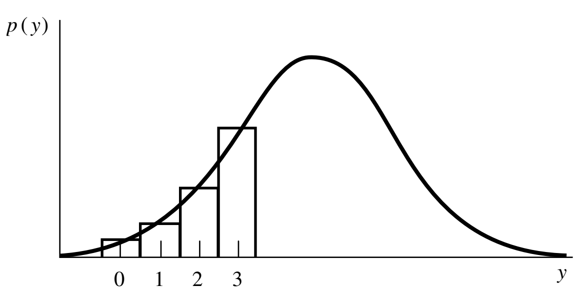
En el Ejercicio 7.70 usted demostrará que un criterio más conveniente pero similar es que la aproximación normal es adecuada si

$$n > 9\left(\frac{\text{el mayor de } p \text{ y } q}{\text{el menor de } p \text{ y } q}\right).$$

Como veremos en el Ejercicio 7.71, para algunos valores de p, en ocasiones se satisface este criterio para valores moderados de n. En especial para valores moderados de n se puede obtener una mejoría considerable en la aproximación mediante un ligero ajuste en las fronteras empleadas en los cálculos. Si vemos el segmento de una distribución binomial graficado en la Figura 7.8, podemos ver lo que ocurre cuando tratamos de calcular una distribución discreta representada por un histograma con una función de densidad continua.

Si deseamos determinar $P(Y \leq 3)$ con el uso de la distribución binomial, podemos calcular el área total de los cuatro rectángulos (arriba de 0, 1, 2 y 3) ilustrados en el histograma binomial (Figura 7.8). Observe que el área total de los rectángulos puede ser aproximada por medio del área bajo la curva normal, la cual incluye algunas áreas que no están en el histograma y excluye la parte del histograma que se encuentre arriba de ella. Si deseamos determinar $P(Y \leq 3)$ al calcular un área bajo la función de densidad, el área bajo la función de densidad localizada a la izquierda de 3.5 da una mejor aproximación que el área localizada a la izquierda de 3.0. El siguiente ejemplo ilustra qué tan cerca está la aproximación normal para un caso en el que pueden hallarse algunas probabilidades binomiales exactas.

FIGURA **7.8** Aproximación normal a la distribución binomial: $n = 10$ y $p = .5$



EJEMPLO **7.11** Suponga que Y tiene una distribución binominal con $n = 25$ y $p = .4$. Encuentre las probabilidades exactas de que $Y \le 8$ y $Y = 8$ y compare éstas con los valores correspondientes determinados con el uso de la aproximación normal.

Solución De la Tabla 1, Apéndice 3, hallamos que

$$P(Y \le 8) = .274$$

y

$$P(Y = 8) = P(Y \le 8) - P(Y \le 7) = .274 - .154 = .120.$$

Como dijimos antes, podemos considerar que Y tiene aproximadamente la misma distribución que W, donde W está distribuida normalmente con $\mu_W = np$ y $\sigma_W^2 = np(1-p)$. Como buscamos $P(Y \le 8)$, vemos el área de la curva normal localizada a la izquierda de 8.5. Así,

$$\begin{aligned} P(Y \le 8) \approx P(W \le 8.5) &= P\left[\frac{W - np}{\sqrt{np(1-p)}} \le \frac{8.5 - 10}{\sqrt{25(.4)(.6)}}\right] \\ &= P(Z \le -.61) = .2709 \end{aligned}$$

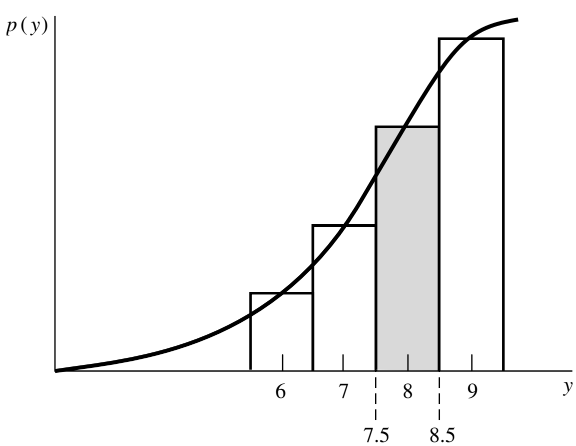
de la Tabla 4, Apéndice 3. Este valor aproximado es cercano al valor exacto para $P(Y \le 8) = .274$, obtenido de las tablas binomiales.

Para determinar la aproximación normal a la probabilidad binomial $p(8)$, calcularemos el área bajo la curva normal entre los puntos 7.5 y 8.5 porque este es el intervalo incluido en la barra del histograma localizada sobre $y = 8$ (véase la Figura 7.9).

Como Y tiene aproximadamente la misma distribución que W, donde W está distribuida normalmente con $\mu_W = np = 25(.4) = 10$ y $\sigma_W^2 = np(1-p) = 25(.4)(.6) = 6$, se deduce que

$$\begin{aligned} P(Y = 8) &\approx P(7.5 \le W \le 8.5) \\ &= P\left(\frac{7.5 - 10}{\sqrt{6}} \le \frac{W - 10}{\sqrt{6}} \le \frac{8.5 - 10}{\sqrt{6}}\right) \\ &= P(-1.02 \le Z \le -.61) = .2709 - .1539 = .1170. \end{aligned}$$

FIGURA **7.9** $P(Y = 8)$ para la distribución binomial del Ejemplo 7.11



Nuevamente vemos que este valor aproximado es muy cercano al valor real, $P(Y = 8) = .120$, calculado antes. ■

Líneas antes, en el ejemplo, empleamos el área bajo una curva normal para calcular $P(Y \leq 8)$ y $P(Y = 8)$ cuando Y tenía una distribución binomial con $n = 25$ y $p = .4$. Para mejorar la aproximación, .5 se sumó al valor máximo de interés (8) cuando usamos la aproximación $P(Y \leq 8) \approx P(W \leq 8.5)$ y W tenía una distribución normal apropiada. De haber estado interesados en calcular $P(Y \geq 6)$, hubiéramos usado $P(Y \geq 6) \approx P(W \geq 5.5)$; esto es, habríamos restado .5 del valor más pequeño de interés (6). El .5 que sumamos al valor máximo de interés (haciéndolo un poco mayor) y sustraemos del valor mínimo de interés (haciéndolo un poco menor), suele recibir el nombre de *corrección de continuidad* asociada con la aproximación normal. La única vez que se usa esta corrección de continuidad en el texto es cuando calculamos una distribución binomial (discreta) con una distribución normal (continua).

Ejercicios

7.64 **Ejercicio Applet** entre en la aplicación *Normal Approximation to Binomial Distribution* (en www.thomsonedu.com/statistics/wackerly). Cuando se inicia la aplicación, ésta exhibe los detalles del Ejemplo 7.11 y la Figura 7.9. Al principio, la pantalla contiene sólo el histograma binomial y el valor exacto (calculado usando la función de probabilidad binomial) para $p(8) = P(Y = 8)$. Arrastre hacia abajo un poco y haga clic en el botón “Toggle Normal Approximation” para recubrir la densidad normal con media 10 y desviación estándar $\sqrt{.6} = 2.449$, las mismas media y desviación estándar que la variable aleatoria binomial Y. Obtendrá una gráfica superior a la de la Figura 7.9.

a ¿Cuántas funciones de probabilidad de masa o densidad se exhiben?

b Introduzca 0 en la caja marcada “Begin” y presione la tecla de Aceptar. ¿Qué probabilidades obtiene?

c Consulte el inciso b. En la línea donde se exhibe la probabilidad normal de aproximación, verá la expresión

$$\text{Normal: } P(-0.5 <= k <= 8.5) = 0.2701.$$

¿Por qué están los .5 en esta expresión?

7.65 **Ejercicio Applet** Suponga que Y tiene una distribución binomial con $n = 5$ y $p = .10$.

a Use la aplicación *Normal Approximation to Binomial Distribution* para hallar valores exactos y aproximados de $P(Y \leq 1)$.

b La aproximación normal no es particularmente buena. ¿Por qué?

7.66 **Ejercicio Applet** Consulte el Ejercicio 7.65. En ese caso, $P(Y \leq 1) = P(|Y - E(Y)| < 1)$. Si $p = .10$, use la aplicación *Normal Approximation to Binomial Distribution* para determinar la n más pequeña para que el valor exacto y la aproximación normal de $P(|Y - E(Y)| < 1)$ difieran en menos de .01.

7.67 **Ejercicio Applet** Suponga que Y tiene una distribución binomial con $p = .20$.

a Use la aplicación *Normal Approximation to Binomial Distribution* para hallar valores exactos y aproximados de $P(Y \leq \mu + 3)$ para $n = 5$, 10, 15 y 20. Para cada tamaño muestral, ponga atención a las formas de los histogramas binomiales y lo cercanas que están las aproximaciones a las probabilidades binomiales exactas.

b Consulte el inciso a. ¿Qué observó acerca de las formas de los histogramas binomiales cuando aumentó el tamaño muestral? ¿Qué observó acerca de las diferencias entre los valores exactos y aproximados para $P(Y \leq \mu + 3)$ a medida que aumentó el tamaño muestral?

c De acuerdo con la regla práctica para lo adecuado de la aproximación normal, ¿qué tan grande debe ser n para que la aproximación sea adecuada? ¿Es esto consistente con lo que observó en los incisos a y b?

7.68 **Ejercicio Applet** En 2004 el estado de Florida resultó afectado por cuatro huracanes de gran intensidad. En 2005 un estudio indicó que en 2004, 48% de las familias en Florida no tenían planes para escapar de un huracán que se aproximaba. Suponga que una muestra aleatoria reciente de 50 familias se seleccionó en Gainesville y que los miembros de 29 de las familias indicaron que tenían un plan de escape en caso de huracán.

a Si los porcentajes estatales de 2004 todavía fueran válidos para las familias recientes de Gainesville, use la aplicación *Normal Approximation to Binomial Distribution* para determinar los valores exacto y aproximado de la probabilidad que 29 o más de las familias muestreadas tengan un plan de escape para el huracán.

b Consulte el inciso a. ¿La aproximación normal es cercana a la probabilidad binomial exacta? Explique por qué.

7.69 Consulte el Ejercicio 7.68.

a Con base en su respuesta al Ejercicio 7.68(a), ¿piensa usted que los porcentajes de 2004 de Florida todavía se aplican a las nuevas familias de Gainesville?

b Sea Y el número de familias de Gainesville que tienen un plan de escape en caso de huracán en una muestra de tamaño 50. Use la aplicación *Normal Approximation to Binomial Distribution* para determinar el valor de b, de modo que $P(Y \geq b)$ sea lo suficientemente pequeña como para llegar a la conclusión de que los porcentajes de Florida de 2004 no se aplican a las nuevas familias de Gainesville.

7.70 En esta sección enunciamos la regla práctica de que la aproximación normal a la distribución binomial es adecuada si $p \pm 3\sqrt{pq/n}$ se encuentra en el intervalo (0, 1), es decir, si

$$0 < p - 3\sqrt{pq/n} \quad \text{y} \quad p + 3\sqrt{pq/n} < 1.$$

a Demuestre que

$$p + 3\sqrt{pq/n} < 1 \text{ si y sólo si } n > 9(p/q).$$

b Demuestre que

$$0 < p - 3\sqrt{pq/n} \text{ si y sólo si } n > 9(q/p).$$

c Combine los resultados de los incisos a y b para obtener que la aproximación normal a la binomial es adecuada si

$$n > 9\left(\frac{p}{q}\right) \quad \text{y} \quad n > 9\left(\frac{q}{p}\right),$$

o bien, de manera equivalente,

$$n > 9\left(\frac{\text{el mayor de } p \text{ y } q}{\text{el menor de } p \text{ y } q}\right)$$

7.71 Consulte el Ejercicio 7.70.

a ¿Para qué valores de n será adecuada la aproximación normal a la distribución binomial si $p = .5$?

b Conteste la pregunta del inciso a si $p = .6, .4, .8, .2, .99$ y $.001$.

7.72 Una máquina se apaga para repararla si una muestra aleatoria de 100 piezas seleccionadas de la producción diaria de la máquina contiene al menos 15% de piezas defectuosas. (Suponga que la producción diaria es un número grande de piezas.) Si en un día determinado la máquina está produciendo sólo 10% de piezas defectuosas, ¿cuál es la probabilidad de que sea apagada? [*Sugerencia:* use la corrección de continuidad .5.]

7.73 Una línea aérea encuentra que 5% de las personas que hacen reservaciones para cierto vuelo no se presentan para el mismo. Si la línea aérea vende 160 boletos para un vuelo con sólo 155 asientos, ¿cuál es la probabilidad de que se disponga de un asiento para cada persona que tiene una reservación y piense volar?

7.74 De acuerdo con una encuesta realizada por la American Bar Association, 1 de cada 410 estadounidenses es abogado, mientras que 1 de cada 64 residentes en Washington, D.C. es abogado.

a Si se seleccionan aleatoriamente 1500 estadounidenses, ¿cuál es la probabilidad aproximada de que la muestra contenga al menos un abogado?

b Si la muestra se selecciona entre los residentes de Washington, D.C., ¿cuál es la probabilidad aproximada de que la muestra contenga más de 30 abogados?

c Si nos encontramos en una esquina en las calles de Washington, D.C. y entrevistamos a las primeras 1000 personas que pasan y 30 dicen que son abogados, ¿esto sugiere que la densidad de abogados que pasan por la esquina es mayor que la densidad dentro de la ciudad? Explique.

7.75 Un entrevistador piensa que 20% de los votantes de cierta zona están a favor de la emisión de bonos. Si 64 votantes se muestrean aleatoriamente de entre el gran número de electores de esta zona, calcule la probabilidad de que la fracción muestreada de votantes que están a favor de la emisión de bonos no difiera en más de .06 de la fracción real.

7.76 **a** Demuestre que la varianza de Y/n, donde Y tiene una distribución binomial con n intentos y una probabilidad de éxito de p, tiene un máximo en $p = .5$ para n fija.

b Se ha de seleccionar una muestra aleatoria de n piezas de un lote grande, observando el número Y de piezas defectuosas. ¿Qué valor de n garantiza que Y/n estará a no más de .1 de la fracción real de piezas defectuosas, con probabilidad de .95?

7.77 El gerente de un supermercado desea obtener información acerca de la proporción de clientes a quienes les disgusta la nueva política de hacer efectivos los cheques. ¿Cuántos clientes debe seleccionar si desea que la fracción muestral se encuentre a no más de .15 de la fracción real, con probabilidad de .98?

7.78 Si el gerente del supermercado (Ejercicio 7.77) muestrea $n = 50$ clientes y si la fracción real de clientes a quienes les disgusta la política es aproximadamente .9, encuentre la probabilidad de que la fracción muestral se encuentre a no más de .15 unidades de la fracción real.

7.79 Suponga que una muestra aleatoria de 25 piezas se selecciona de la máquina del Ejercicio 7.72. Si la máquina produce 10% de piezas defectuosas, encuentre la probabilidad de que la muestra contenga al menos dos piezas defectuosas, usando los siguientes métodos:

a La aproximación normal a la binomial.

b Las tablas binomiales exactas.

7.80 La edad promedio de los residentes de Estados Unidos es 31 años. Si se entrevista a 100 residentes seleccionados aleatoriamente, ¿cuál es la probabilidad aproximada de que al menos 60 tengan menos de 31 años de edad?

7.81 Un plan de muestreo de aceptación de un lote especifica que 50 piezas sean seleccionadas aleatoriamente y que el lote sea aceptado si no más de 5 de las piezas elegidas no se ajustan a las especificaciones.

a ¿Cuál es la probabilidad aproximada de que un lote sea aceptado si la verdadera proporción de piezas que no se ajustan a especificaciones en el lote es .10?

b Conteste la pregunta del inciso a si la verdadera proporción de piezas que no se ajustan a especificaciones en el lote es .20 y .30.

7.82 La calidad de discos de computadora se mide por el número de pulsos faltantes. La marca X es tal que 80% de los discos no tienen pulsos faltantes. Si se inspeccionan 100 discos de la marca X, ¿cuál es la probabilidad de que 15 o más contengan pulsos faltantes?

7.83 **Ejercicio Applet** Los vehículos que entran a un crucero desde el este tienen igual probabilidad de dar vuelta a la izquierda, a la derecha o continuar de frente. Si 50 vehículos entran a este crucero desde el este, use la aplicación *Normal Approximation to Binomial Distribution* para determinar las probabilidades exactas y aproximadas de que

a 15 o menos den vuelta a la derecha.

b Al menos dos tercios de la muestra den vuelta.

7.84 Así como la diferencia entre dos medias muestrales está distribuida normalmente para muestras grandes, la diferencia entre dos proporciones muestrales está distribuida de la misma manera. Es decir, si Y_1 y Y_2 son variables aleatorias binomiales independientes con parámetros (n_1, p_1) y (n_2, p_2), respectivamente, entonces $(Y_1/n_1) - (Y_2/n_2)$ está distribuida normalmente en forma aproximada para grandes valores de n_1 y n_2.

a Encuentre $E\left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right)$.

b Encuentre $V\left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right)$.

7.85 Para verificar la abundancia relativa de cierta especie de peces en dos lagos, se toman $n = 50$ observaciones relacionadas con los resultados de la captura en cada uno de los lagos. Para cada observación, el experimentador sólo registra si la especie deseada estaba presente en la trampa. La experiencia del pasado ha demostrado que esta especie aparece en trampas del lago A aproximadamente 10% del tiempo y en trampas del lago B, alrededor de 20% del tiempo. Use estos resultados para aproximar la probabilidad de que la diferencia entre las proporciones muestrales sea de no más de .1 de la diferencia entre las proporciones reales.

7.86 Un auditor muestrea 100 de los comprobantes de viaje de una empresa para averiguar qué porcentaje de entre el conjunto total de comprobantes está mal documentado. ¿Cuál es la probabilidad aproximada de que más de 30% de los comprobantes muestreados estén mal documentados si, de hecho, sólo 20% de todos los comprobantes están mal documentados? Si usted fuera el auditor y observa que más de 30% está mal documentado, ¿qué concluiría acerca de lo dicho por la empresa de que sólo 20% sufría de mala documentación? ¿Por qué?

7.87 Los tiempos para procesar pedidos en el mostrador de servicio de una farmacia están distribuidos exponencialmente con media de 10 minutos. Si 100 clientes pasan al mostrador en un periodo de 2 días, ¿cuál es la probabilidad de que al menos la mitad de ellos necesite esperar más de 10 minutos?

7.6 Resumen

Para hacer inferencias acerca de parámetros poblacionales, necesitamos conocer las distribuciones de probabilidad para ciertos *estadísticos*, funciones de las variables aleatorias observables de la muestra (o muestras). Estas distribuciones de probabilidad dan modelos para el

Tabla 7.2 Procedimientos *R* (y *S*-Plus) que dan probabilidades y percentiles para distribuciones normales, χ^2, *t* y *F*

Distribución	$P(Y \leq y_0)$	p-ésimo cuantil ϕ_p tal que $P(Y \leq \phi_p) = p$
Normal (μ,σ)	pnorm(y_0, μ, σ)	qnorm(p, μ, σ)
χ^2 con ν grados de libertad	pchisq(y_0, ν)	qchisq(p, ν)
t con ν grados de libertad	pt(y_0, ν)	qt(p, ν)
F con ν_1 gl en el numerador, ν_2 gl en el denominador	pf(y_0, ν_1, ν_2)	qf(p, ν_1, ν_2)

comportamiento de frecuencia relativa de los estadísticos en muestreo repetido; en consecuencia, se conocen como *distribuciones muestrales.* Hemos visto que las distribuciones normales, χ^2, t y F dan modelos para las distribuciones de muestreo de estadísticos empleados para hacer inferencias acerca de los parámetros asociados con distribuciones normales. Para mayor comodidad, la Tabla 7.2 contiene un resumen de los comandos *R* (o *S*-Plus) que proporcionan probabilidades y cuantiles asociados con estas distribuciones.

Cuando el tamaño muestral es grande, la media muestral $\overline{Y}$ posee una distribución aproximadamente normal si la muestra aleatoria se toma de *cualquier* distribución con media finita μ y una varianza finita σ^2. Este resultado, conocido como el *teorema del límite central*, también proporciona la justificación para calcular probabilidades binomiales con probabilidades correspondientes asociadas con la distribución normal.

Las distribuciones muestrales expuestas en este capítulo se usarán en los procedimientos para hacer inferencias que se presentan en capítulos subsecuentes.

Bibliografía y lecturas adicionales

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2nd ed. Pacific Grove, Calif.: Duxbury.

Hoel, P. G. 1984. *Introduction to Mathematical Statistics,* 5th ed. New York: Wiley.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics,* 6th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics,* 3d ed. New York: McGraw-Hill.

Parzen, E. 1992. *Modern Probability Theory and Its Applications*. New York: Wiley-Interscience.

Ejercicios complementarios

7.88 La eficiencia (en lúmenes por watt) de los focos de cierto tipo tiene una media poblacional de 9.5 y desviación estándar de .5, de acuerdo con especificaciones de producción. Las especificaciones para un cuarto en el que ocho de estos focos se han de instalar exigen que el promedio de eficiencia de los mismos sea mayor que 10. Encuentre la probabilidad de que se satisfaga esta especificación para el cuarto, suponiendo que las mediciones de eficiencia están distribuidas normalmente.

7.89 Consulte el Ejercicio 7.88. ¿Cuál debe ser la eficiencia media por foco si debe satisfacerse la especificación para el cuarto con una probabilidad de aproximadamente .80? (Suponga que la varianza de las mediciones de eficiencia continúa en .5.)

7.90 Briggs y King desarrollaron la técnica de trasplante nuclear, en el que el núcleo de una célula de una de las etapas finales de desarrollo de un embrión se trasplanta en un cigoto (un huevo fertilizado de una sola célula), para ver si el núcleo puede soportar un desarrollo normal. Si .65 es la probabilidad de que un solo trasplante de la etapa temprana de gástrula sea exitosa, ¿cuál es la probabilidad de que más de 70 trasplantes de entre 100 sea exitosa?

7.91 Un distribuidor minorista vende tres marcas de automóviles. Para la marca A, su utilidad X por venta, está distribuida normalmente con parámetros (μ_1, σ_1^2); para la marca B su utilidad Y por venta está distribuida normalmente con parámetros (μ_2, σ_2^2); para la marca C, su utilidad W por venta está distribuida normalmente con parámetros (μ_3, σ_3^2). Para un año, dos quintas partes de las ventas del distribuidor son de la marca A, un quinto de la marca B y los dos quintos restantes de la marca C. Si contamos con datos sobre las utilidades por ventas n_1, n_2 y n_3 de las marcas A, B y C, respectivamente, la cantidad $U = .4\overline{X} + .2\overline{Y} + .4\overline{W}$ calculará el verdadero promedio de utilidad por ventas para el año. Encuentre la media, la varianza y la función de densidad de probabilidad para U. Suponga que X, Y y W son independientes.

7.92 De cada una de dos poblaciones normales con medias idénticas y con desviaciones estándar de 6.40 y 7.20, se toman muestras aleatorias independientes de 64 observaciones. Encuentre la probabilidad de que la diferencia entre las medias de las muestras exceda de .6 en valor absoluto.

7.93 Si Y tiene una distribución exponencial con media θ, demuestre que $U = 2Y/\theta$ tiene una distribución χ^2 con 2 grados de libertad.

7.94 Un supervisor de planta está interesado en asignar presupuesto a los costos semanales de reparaciones para cierto tipo de máquina. Los registros de años pasados indican que estos costos de reparación tienen una distribución exponencial con media de 20 para cada máquina estudiada. Denote con $Y_1, Y_2, \ldots, Y_5$ los costos de reparación para cinco de estas máquinas durante la semana siguiente. Encuentre un número c tal que $P\left(\sum_{i=1}^{5} Y_i > c\right) = .0.5$, suponiendo que las máquinas operan de manera independiente. [*Sugerencia:* use el resultado dado en el Ejercicio 7.93.]

7.95 El *coeficiente de variación* (CV) para una muestra de valores $Y_1, Y_2, \ldots, Y_n$ está definido por

$$\text{CV} = S/\overline{Y}.$$

Esta cantidad, que suministra la desviación estándar como una proporción de la media, en ocasiones es informativa. Por ejemplo, el valor $S = 10$ tiene poco significado a menos que podamos compararlo con algo más. Si se observa que S es igual a 10 y $\overline{Y}$ es igual a 1000, la cantidad de variación es pequeña en comparación con el tamaño de la media. No obstante, si se observa que S es 10 y que $\overline{Y}$ es 5, la variación es bastante grande con respecto al tamaño de la media. Si estuviéramos estudiando la precisión (variación en mediciones repetidas) de un instrumento de medición, el primer caso (CV = 10/1000) podría dar una precisión aceptable, pero el segundo caso (CV = 2) sería inaceptable.

Denote con $Y_1, Y_2, \ldots, Y_{10}$ una muestra aleatoria de tamaño 10 tomada de una distribución normal con media 0 y varianza σ^2. Use los siguientes pasos para determinar el número c tal que

$$P\left(-c \le \frac{S}{\overline{Y}} \le c\right) = .95.$$

a Use el resultado del Ejercicio 7.33 para determinar la distribución de $(10)\,\overline{Y}^2/S^2$.

b Use el resultado del Ejercicio 7.29 para determinar la distribución de $S^2/[(10)\overline{Y}^2]$.

c Use la respuesta al inciso b para determinar la constante c.

7.96 Suponga que $Y_1, Y_2, \ldots, Y_{40}$ denota una muestra aleatoria de mediciones de la proporción de impurezas en muestras de mineral de hierro. Cada variable Y_i tiene una función de densidad de probabilidad dada por

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

El mineral ha de ser rechazado por el comprador potencial si $\overline{Y}$ excede de .7. Encuentre $P(\overline{Y} > .7)$ para la muestra de tamaño 40.

***7.97** Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes y con distribución χ^2, cada una con grado de libertad 1. Defina Y como

$$Y = \sum_{i=1}^{n} X_i.$$

Se deduce del Ejercicio 6.59 que Y tiene una distribución χ^2 con n grados de libertad.

a Use la representación anterior de Y como la suma de las X para demostrar que $Z = (Y - n)/\sqrt{2n}$ tiene una distribución normal estándar asintótica.

b Una máquina en una fábrica de equipo pesado produce varillas de acero de longitud Y, donde Y es una variable aleatoria distribuida normalmente con media de 6 pulgadas y varianza .2. El costo C de reparar una varilla que no mida exactamente 6 pulgadas de largo es proporcional al cuadrado del error y está dado, en dólares, por $C = 4(Y - \mu)^2$. Si 50 varillas con longitudes independientes se producen en un día determinado, aproxime la probabilidad de que el costo total de reparaciones para ese día exceda de \$48.

***7.98** Suponga que T se precisa como en la Definición 7.2.

a Si W tiene un valor fijo w, entonces T está dada por Z/c, donde $c = \sqrt{w/\nu}$. Use esta idea para determinar la densidad condicional de T para una $W = w$ fija.

b Encuentre la densidad conjunta de T y W, $f(t, w)$, usando $f(t, w) = f(t|w)f(w)$.

c Integre sobre w para demostrar que

$$f(t) = \left\{ \frac{\Gamma[(\nu + 1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \right\} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \qquad -\infty < t < \infty.$$

***7.99** Suponga que F se precisa como en la Definición 7.3.

a Si W_2 está fija en w_2, entonces $F = W_1/c$, donde $c = w_2\nu_1/\nu_2$. Encuentre la densidad condicional de F para $W_2 = w_2$ fija.

b Encuentre la densidad conjunta de F y W_2.

c Integre sobre w_2 para demostrar que la función de densidad de probabilidad de F, $g(y)$ por ejemplo, está dada por

$$g(y) = \frac{\Gamma[(\nu_1 + \nu_2)/2](\nu_1/\nu_2)^{\nu_1/2}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} y^{(\nu_1/2)-1} \left(1 + \frac{\nu_1 y}{\nu_2}\right)^{-(\nu_1+\nu_2)/2}, \qquad 0 < y < \infty.$$

***7.100** Sea X que tiene una distribución de Poisson con parámetro λ.

a Demuestre que la función generadora de momento de $Y = (X - \lambda)/\sqrt{\lambda}$ está dada por

$$m_Y(t) = \exp(\lambda e^{t/\sqrt{\lambda}} - \sqrt{\lambda}t - \lambda).$$

b Use la expansión

$$e^{t/\sqrt{\lambda}} = \sum_{i=0}^{\infty} \frac{[t/\sqrt{\lambda}]^i}{i!}$$

para demostrar que

$$\lim_{\lambda \to \infty} m_Y(t) = e^{t^2/2}.$$

c Use el Teorema 7.5 para demostrar que la función de distribución de Y converge hacia una función de distribución normal estándar cuando $\lambda \to \infty$.

***7.101** En el interés de controlar la contaminación, un experimentador desea contar el número de bacterias en un pequeño volumen de agua. Denote con X la cantidad de bacterias por centímetro cúbico de agua y suponga que X tiene una distribución de probabilidad de Poisson con media $\lambda = 100$. Si la contaminación permisible en un suministro de agua es de 110 bacterias por centímetro cúbico, calcule la probabilidad de que X sea a lo sumo de 110. [*Sugerencia:* use el resultado del Ejercicio 7.100(c).]

***7.102** Se supone que Y, el número de accidentes por año en cierto crucero, tiene una distribución de Poisson. En años recientes, un promedio de 36 accidentes por año han ocurrido en este crucero. Si el número de accidentes por año es al menos de 45, un crucero puede tener requisitos para ser rediseñado dentro de un programa de emergencia establecido por el Estado. Calcule la probabilidad de que el crucero en cuestión entre en el programa de emergencia a fines del año próximo.

***7.103** Un experimentador está comparando dos métodos para remover colonias de bacterias contenidas en carnes frías. Después de tratar algunas muestras mediante el método A y otras muestras con el método B, el experimentador selecciona una submuestra de 2 centímetros cúbicos de cada muestra y hace la suma de la colonia de bacterias de las submuestras. Denote con X la cantidad total para las submuestras tratadas con el método A y denote con Y la cantidad total para las tratadas con el método B. Suponga que X y Y son variables aleatorias de Poisson independientes, con medias λ_1 y λ_2, respectivamente. Si X excede de Y en más de 10, el método B se juzgará mejor que el A. Suponga que, de hecho, $\lambda_1 = \lambda_2 = 50$. Encuentre la probabilidad aproximada de que el método B sea considerado mejor que el método A.

***7.104** Sea Y_n una variable aleatoria binomial con n intentos y con probabilidad p de éxito. Suponga que n tiende al infinito y que p tiende a cero en forma tal que np continúa fija en $np = \lambda$. Use el resultado del Teorema 7.5 para demostrar que la distribución de Y_n converge hacia una distribución de Poisson con media de λ.

***7.105** Si la probabilidad de que una persona sufra de una reacción adversa por un medicamento es .001, use el resultado del Ejercicio 7.104 para calcular la probabilidad de que 2 o más personas sufran una reacción adversa si el medicamento se administra a 100 personas.

CAPÍTULO 8

Estimación

8.1 Introducción

Como lo establecimos en el Capítulo 1, el propósito de la estadística es usar la información contenida en una muestra para hacer inferencias acerca de la población de la cual se toma la muestra. Debido a que las poblaciones están caracterizadas por medidas descriptivas numéricas llamadas *parámetros*, el objetivo de muchas investigaciones estadísticas es calcular el valor de uno o más parámetros relevantes. Como veremos, las distribuciones muestrales obtenidas en el Capítulo 7 desempeñan un importante papel en el desarrollo de los procedimientos de estimación que son el objetivo de este capítulo.

La estimación tiene muchas aplicaciones prácticas. Por ejemplo, un fabricante de máquinas lavadoras podría estar interesado en estimar la proporción p de lavadoras que esperaría que fallen antes de la expiración de la garantía de un año. Otros parámetros poblacionales importantes son la media poblacional, la varianza y la desviación estándar. Por ejemplo, podríamos estimar la media del tiempo de espera μ en una caja registradora del supermercado o la desviación estándar del error de medición σ de un instrumento electrónico. Para simplificar nuestra terminología, al parámetro de interés le llamaremos *parámetro objetivo* en el experimento.

Suponga que deseamos estimar la cantidad promedio μ de mercurio que un proceso recién inventado puede eliminar de 1 onza de mineral obtenido de un lugar geográfico determinado. Podríamos dar nuestra estimación o cálculo en dos formas distintas. Primero, podríamos usar un solo número, por ejemplo .13 onzas, que consideramos es cercano a la media poblacional desconocida μ. Este tipo de estimación se llama *estimación puntual* porque un solo valor o punto constituye la estimación de μ. En segundo término podríamos decir que μ está entre dos números, por ejemplo entre .07 y .19 onzas. En este segundo procedimiento de estimación los dos valores se pueden utilizar para construir un intervalo (.07, .19) que tiene la intención de encerrar el parámetro de interés; entonces, la estimación se denomina *estimación de intervalo.*

La información de la muestra se puede emplear para calcular el valor de una estimación puntual, una estimación de intervalo o ambas. En cualquier caso, la estimación real se logra con el uso de un *estimador* del parámetro objetivo.

DEFINICIÓN **8.1**

Un *estimador* es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.

Por ejemplo, la media muestral

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

es un posible estimador puntual de la media poblacional μ. Claramente, la expresión para $\overline{Y}$ es una regla y una fórmula. Nos indica que sumemos las observaciones muestrales y dividamos entre el tamaño muestral n.

Un investigador que necesite una estimación de intervalo de un parámetro debe usar los datos muestrales para calcular dos valores, escogidos de tal modo que el intervalo que formen incluya el parámetro objetivo con una probabilidad específica. En las siguientes secciones se darán ejemplos de estimadores por intervalos.

Muchos estimadores diferentes (reglas de estimación) pueden obtenerse para el mismo parámetro poblacional. Esto no debe ser sorprendente. Diez ingenieros, cada uno de ellos asignado a estimar el costo de un gran trabajo de construcción, podrían usar métodos diferentes de estimación y por tanto llegar a diferentes evaluaciones del costo total. Estos ingenieros, llamados *estimadores* en la industria de la construcción, basan sus estimaciones en lineamientos fijos especificados y en la intuición. Cada estimador representa una regla subjetiva humana única para obtener una sola estimación. Esto nos lleva a un punto más importante: algunos estimadores se consideran *buenos* y otros *malos*. La administración de una empresa constructora debe definir entre *bueno* y *malo* cuando en lo relacionado con la estimación del costo de un trabajo. ¿Cómo podemos establecer criterios de bondad para comparar estimadores estadísticos? Las siguientes secciones contienen algunas respuestas a esta pregunta.

8.2 Sesgo y error cuadrático medio de estimadores puntuales

La estimación puntual es similar, en muchos aspectos, a disparar a un blanco con un revólver. El estimador, que genera estimaciones, es análogo al revólver; una estimación particular es comparable a un tiro; y el parámetro de interés corresponde al centro del blanco o diana. Extraer una sola muestra de una población y usarla para calcular una estimación del valor del parámetro equivale a disparar un solo tiro al centro del blanco.

Suponga que un hombre dispara un solo tiro y acierta en el centro del blanco. ¿Concluimos que es un excelente tirador? ¿Se atrevería a sostener el blanco cuando se haga un segundo tiro? Obviamente, no podríamos afirmar que el hombre es un experto tirador si nos basamos en tan pequeña cantidad de evidencia. Por otra parte, si 100 tiros en sucesión aciertan en el blanco, podríamos tener confianza suficiente en el tirador y considerar sostener el blanco para el siguiente tiro si la compensación es adecuada. La cuestión es que no podemos evaluar la bondad de un procedimiento de estimación puntual con base en el valor de una sola estimación; más bien, debemos observar los resultados cuando el procedimiento de estimación se usa en innumerables veces. Como las estimaciones son números, evaluamos la bondad del estimador puntual al construir una distribución de frecuencia de los valores de las estimaciones obtenidas en muestreo repetido y observar cómo se agrupa esta distribución alrededor del parámetro objetivo.

Suponga que deseamos especificar una estimación puntual para un parámetro poblacional al que llamaremos θ. El estimador de θ estará indicado por el símbolo $\hat{\theta}$, que se lee como "θ sombrero". El "sombrero" indica que estamos estimando el parámetro que está inmediatamente bajo él. Con el ejemplo del disparo de revólver en mente, podemos decir que es altamente deseable que la distribución de estimaciones —o bien, en forma más apropiada, la distribución muestral de estimaciones— se agrupe alrededor del parámetro objetivo como se muestra en la Figura 8.1. En otras palabras, quisiéramos que la media o valor esperado de la distribución de estimaciones fuera igual al parámetro estimado; esto es, $E(\hat{\theta}) = \theta$. Se dice que los estimadores puntuales que satisfacen esta propiedad son *insesgados*. La distribución muestral para un estimador puntual sesgado positivamente, para el que $E(\hat{\theta}) > \theta$, se muestra en la Figura 8.2.

FIGURA **8.1** Distribución de estimaciones

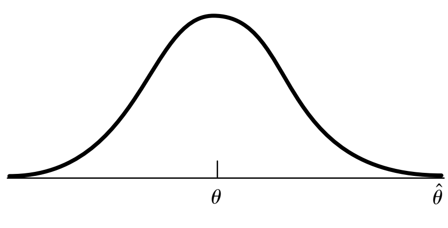
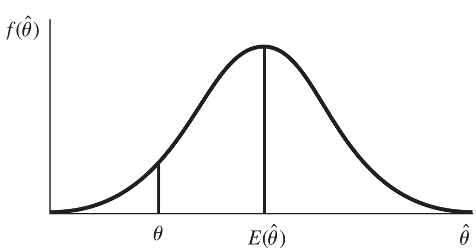


FIGURA **8.2** Distribución muestral para un estimador sesgado positivamente



DEFINICIÓN **8.2**

Si $\hat{\theta}$ es un estimador puntual de un parámetro θ, entonces $\hat{\theta}$ es un *estimador insesgado* si $E(\hat{\theta}) = \theta$. Si $E(\hat{\theta}) \neq \theta$, se dice que $\hat{\theta}$ está *sesgado*.

DEFINICIÓN **8.3**

El *sesgo* de un estimador puntual $\hat{\theta}$ está dado por $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

La Figura 8.3 muestra dos posibles distribuciones muestrales para los estimadores puntuales insesgados de un parámetro objetivo θ. Preferiríamos que nuestro estimador tuviera el tipo de distribución indicado en la Figura 8.3(b) porque una varianza pequeña garantiza que, en un muestreo repetido, una fracción más alta de valores de $\hat{\theta}_2$ estará "cerca" de θ. Por consiguiente, además de preferir un estimador insesgado, necesitamos que la varianza de la distribución del estimador $V(\hat{\theta})$ sea lo más pequeña posible. Dados dos estimadores insesgados de un parámetro θ seleccionaríamos el estimador con la menor varianza mientras, todo lo demás permanece igual.

Más que usar el sesgo y la varianza de un estimador puntual para caracterizar su bondad, podríamos emplear $E[(\hat{\theta} - \theta)^2]$, el promedio del cuadrado de la distancia entre el estimador y su parámetro objetivo.

DEFINICIÓN **8.4**

El *error cuadrático medio* de un estimador puntual $\hat{\theta}$ es

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

El error cuadrático medio de un estimador $\hat{\theta}$, MSE $(\hat{\theta})$, es una función de su varianza y su sesgo. Si $B(\hat{\theta})$ representa el sesgo del estimador $\hat{\theta}$, se puede demostrar que

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2.$$

Dejaremos la demostración de este resultado como Ejercicio 8.1.

En esta sección hemos definido propiedades deseables de los estimadores puntuales. En particular, a menudo buscamos estimadores insesgados con varianzas relativamente pequeñas. En la siguiente sección estudiaremos algunos estimadores puntuales insesgados comunes y útiles.

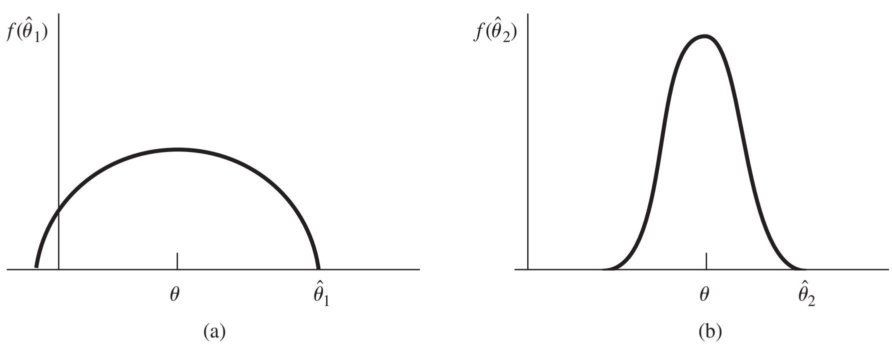


FIGURA **8.3** Distribuciones muestrales para dos estimadores insesgados: (a) estimador con variación grande; (b) estimador con variación pequeña

Ejercicios

8.1 Usando la identidad

$$(\hat{\theta} - \theta) = [\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta] = [\hat{\theta} - E(\hat{\theta})] + B(\hat{\theta}),$$

demuestre que

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + (B(\hat{\theta}))^2.$$

8.2 **a** Si $\hat{\theta}$ es un estimador insesgado para θ, ¿cuál es $B(\hat{\theta})$?

b Si $B(\hat{\theta}) = 5$, ¿cuál es $E(\hat{\theta})$?

8.3 Suponga que $\hat{\theta}$ es un estimador para un parámetro θ y $E(\hat{\theta}) = a\theta + b$ para algunas constantes diferentes de cero a y b.

a En términos de a, b, y θ, ¿cuál es $B(\hat{\theta})$?

b Encuentre una función de $\hat{\theta}$, por ejemplo $\hat{\theta}$,* que es un estimador insesgado para θ.

8.4 Consulte el Ejercicio 8.1.

a Si $\hat{\theta}$ es un estimador insesgado para θ, cómo se compara $\text{MSE}(\hat{\theta})$ con $V(\hat{\theta})$?

b Si $\hat{\theta}$ es un estimador insesgado para θ, cómo se compara $\text{MSE}(\hat{\theta})$ con $V(\hat{\theta})$?

8.5 Consulte los Ejercicios 8.1 y considere el estimador insesgado $\hat{\theta}^*$ que usted propuso en el Ejercicio 8.3.

a Exprese $\text{MSE}(\hat{\theta}^*)$ como función de $V(\hat{\theta})$.

b Dé un ejemplo de un valor para a para el cual $\text{MSE}(\hat{\theta}^*) < \text{MSE}(\hat{\theta})$.

c Dé un ejemplo de valores para a y b para los cuales $\text{MSE}(\hat{\theta}^*) > \text{MSE}(\hat{\theta})$.

8.6 Suponga que $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $V(\hat{\theta}_1) = \sigma_1^2$, y $V(\hat{\theta}_2) = \sigma_2^2$. Considere el estimador $\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$.

a Demuestre que $\hat{\theta}_3$ es un estimador insesgado para θ.

b Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes, ¿cómo debe escogerse la constante a para minimizar la varianza de $\hat{\theta}_3$?

8.7 Considere la situación descrita en el Ejercicio 8.6. ¿Cómo debe elegirse la constante a para minimizar la varianza de $\hat{\theta}_3$, si $\hat{\theta}_1$ y $\hat{\theta}_2$ no son independientes pero son tales que $\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = c \neq 0$?

8.8 Suponga que Y_1, Y_2, Y_3 denotan una muestra aleatoria de una distribución exponencial con función de densidad

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta}\right) e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Considere los siguientes cinco estimadores de θ:

$$\hat{\theta}_1 = Y_1, \qquad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \qquad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \qquad \hat{\theta}_4 = \text{mín}(Y_1, Y_2, Y_3), \qquad \hat{\theta}_5 = \overline{Y}.$$

a ¿Cuáles de estos estimadores son insesgados?

b Entre los estimadores insesgados, ¿cuál tiene la varianza más pequeña?

8.9 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una población con función de densidad de probabilidad

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta + 1}\right) e^{-y/(\theta+1)}, & y > 0, \theta > -1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Sugiera un estadístico apropiado para usarlo como estimador insesgado para θ. [*Sugerencia:* considere $\overline{Y}$.]

8.10 El número de descomposturas por semana para un tipo de minicomputadora es una variable aleatoria Y con una distribución de Poisson y media λ. Existe una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de observaciones del número semanal de descomposturas.

a Sugiera un estimador insesgado para λ.

b El costo semanal de reparar estas descomposturas es $C = 3Y + Y^2$. Demuestre que $E(C) = 4\lambda + \lambda^2$.

c Encuentre una función de $Y_1, Y_2, \ldots, Y_n$ que sea un estimador insesgado de $E(C)$. [*Sugerencia:* use lo que sepa acerca de $\overline{Y}$ y $(\overline{Y})^2$.]

8.11 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población con media 3. Suponga que $\hat{\theta}_2$ es un estimador insesgado de $E(Y^2)$ y que $\hat{\theta}_3$ es un estimador insesgado de $E(Y^3)$. Proponga un estimador insesgado para el tercer momento central de la distribución subyacente.

8.12 La lectura en un voltímetro conectado a un circuito de prueba está distribuida uniformemente en el intervalo $(\theta, \theta + 1)$, donde θ es el valor desconocido del voltaje real del circuito. Suponga que $Y_1, Y_2, \ldots, Y_n$ denota una muestra aleatoria de esas lecturas.

a Demuestre que $\overline{Y}$ es un estimador sesgado de θ y calcule el sesgo.

b Encuentre una función de $\overline{Y}$ que sea un estimador insesgado de θ.

c Encuentre $\text{MSE}(\overline{Y})$ cuando $\overline{Y}$ se use como estimador de θ.

8.13 Hemos visto que si Y tiene una distribución binomial con parámetros n y p, entonces Y/n es un estimador insesgado de p. Para calcular la varianza de Y, por lo general usamos $n(Y/n)(1 - Y/n)$.

a Demuestre que el estimador sugerido es un estimador sesgado de $V(Y)$.

b Modifique ligeramente $n(Y/n)(1 - Y/n)$ para formar un estimador insesgado de $V(Y)$.

8.14 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población cuya densidad está dada por

$$f(y) = \begin{cases} \alpha y^{\alpha-1}/\theta^{\alpha}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\alpha > 0$ es un valor fijo conocido, pero θ no se conoce. (Ésta es la familia de distribuciones de potencia introducidas en el ejercicio 6.17.) Considere el estimador $\hat{\theta} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$.

a Demuestre que $\hat{\theta}$ es un estimador sesgado de $\hat{\theta}$.

b Determine un multiplo de $\hat{\theta}$ que constituya un estimador de $\hat{\theta}$.

c Deduzca $\text{MSE}(\hat{\theta})$

8.15 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población cuya densidad está dada por

$$f(y) = \begin{cases} 3\beta^3 y^{-4}, & \beta \le y, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\beta > 0$ es desconocido. (Ésta es una de las distribuciones de Pareto introducidas en el Ejercicio 6.18.) Considere el estimador $\hat{\beta} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$.

a Deduzca el sesgo del estimador $\hat{\beta}$.

b Deduzca $\text{MSE}(\hat{\beta})$.

***8.16** Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con parámetros μ y σ^2.[1]

a Demuestre que $S = \sqrt{S^2}$ es un estimador sesgado de σ. [*Sugerencia:* recuerde la distribución de $(n-1)S^2/\sigma^2$ y el resultado dado en el Ejercicio 4.112.]

b Ajuste S para formar un estimador insesgado de σ.

c Encuentre un estimador insesgado de $\mu - z_\alpha\sigma$, el punto que corta un área de cola inferior de α bajo esta curva normal.

8.17 Si Y tiene una distribución binomial con parámetros n y p, entonces $\hat{p}_1 = Y/n$ es un estimador insesgado de p. Otro estimador de p es $\hat{p}_2 = (Y+1)/(n+2)$.

a Deduzca el sesgo de $\hat{p}_2$.

b Deduzca $\text{MSE}(\hat{p}_1)$ y $\text{MSE}(\hat{p}_2)$.

c ¿Para qué valores de p es $\text{MSE}(\hat{p}_1) < \text{MSE}(\hat{p}_2)$?

8.18 Si $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria de tamaño n de una población con una distribución uniforme en el intervalo $(0, \theta)$, considere a $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ el estadístico de orden más bajo. Aplique los métodos de la Sección 6.7 para deducir $E(Y_{(1)})$. Encuentre un múltiplo de $Y_{(1)}$ que sea un estimador insesgado para θ.

8.19 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una población con una distribución exponencial cuya densidad está dada por

$$f(y) = \begin{cases} (1/\theta)\, e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Si $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ denota el estadístico de orden más bajo, demuestre que $\hat{\theta} = nY_{(1)}$ es un estimador insesgado para θ y encuentre $\text{MSE}(\hat{\theta})$. [*Sugerencia:* recuerde los resultados del Ejercicio 6.81.]

***8.20** Suponga que Y_1, Y_2, Y_3, Y_4 denotan una muestra aleatoria de tamaño 4 de una población con una distribución exponencial cuya densidad está dada por

$$f(y) \begin{cases} (1/\theta)e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Sea $X = \sqrt{Y_1Y_2}$. Encuentre un múltiplo de X que sea un estimador insesgado para θ. [*Sugerencia:* use su conocimiento de la distribución gamma y el hecho de que $\Gamma(1/2) = \sqrt{\pi}$ para hallar $E(\sqrt{Y_1})$. Recuerde que las variables Y_i son independientes.]

b Sea $W = \sqrt{Y_1Y_2Y_3Y_4}$. Encuentre un múltiplo de W que sea un estimador insesgado para θ^2. [Recuerde la sugerencia para el inciso a.]

8.3 Algunos estimadores puntuales insesgados comunes

Algunos métodos formales para obtener estimadores puntuales de parámetros objetivo se presentan en el Capítulo 9. En esta sección nos concentramos en algunos estimadores que ameritan consideración con base en la intuición. Por ejemplo, parece natural usar la media muestral $\overline{Y}$ para estimar la media poblacional μ y usar la proporción muestral $\hat{p} = Y/n$ para estimar

1 Los ejercicios precedidos por un asterisco son opcionales.

un parámetro binomial p. Si una inferencia está basada en muestras aleatorias independientes de n_1 y n_2 observaciones seleccionadas de dos poblaciones diferentes, ¿cómo estimaríamos la diferencia entre medias $(\mu_1 - \mu_2)$ o la diferencia en dos parámetros binomiales, $(p_1 - p_2)$? De nuevo, nuestra intuición sugiere utilizar los estimadores puntuales $(\overline{Y}_1 - \overline{Y}_2)$, la diferencia en las medias muestrales, para estimar $(\mu_1 - \mu_2)$ y usar $(\hat{p}_1 - \hat{p}_2)$, la diferencia en las proporciones muestrales, para estimar $(p_1 - p_2)$.

Como los cuatro estimadores $\overline{Y}$, $\hat{p}$, $(\overline{Y}_1 - \overline{Y}_2)$ y $(\hat{p}_1 - \hat{p}_2)$ son funciones de las variables aleatorias observadas en muestras, podemos hallar sus valores y varianzas esperados con el uso de teoremas de esperanza de las Secciones 5.6-5.8. La desviación estándar de cada uno de los estimadores es simplemente la raíz cuadrada de la varianza respectiva. Ese esfuerzo demostraría que, cuando se ha empleado un muestreo aleatorio, los cuatro estimadores puntuales son insesgados y que poseen las desviaciones estándar mostradas en la Tabla 8.1. Para facilitar la exposición, usamos la notación $\sigma^2_{\hat{\theta}}$ para denotar la varianza de la distribución muestral del estimador $\hat{\theta}$. La desviación estándar de la distribución muestral del estimador $\hat{\theta}$, $\hat{\theta}$, $\sigma_{\hat{\theta}} = \sqrt{\sigma^2_{\hat{\theta}}}$, suele recibir el nombre de *error estándar* del estimador $\hat{\theta}$.

En el Capítulo 5 dedujimos gran parte de la información requerida para la Tabla 8.1. En particular, determinamos las medias y varianzas de $\overline{Y}$ y $\hat{p}$ en los Ejemplos 5.27 y 5.28, respectivamente. Si las muestras aleatorias son independientes, estos resultados y el Teorema 5.12 implican que

$$E(\overline{Y}_1 - \overline{Y}_2) = E(\overline{Y}_1) - E(\overline{Y}_2) = \mu_1 - \mu_2,$$
$$V(\overline{Y}_1 - \overline{Y}_2) = V(\overline{Y}_1) + V(\overline{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

El valor esperado y el error estándar de $(\hat{p}_1 - \hat{p}_2)$, mostrados en la Tabla 8.1, se pueden obtener de un modo semejante.

Tabla 8.1 Valores esperados y errores estándar de algunos estimadores puntuales comunes

Parámetro objetivo θ	Tamaño(s) muestral(es)	Estimador puntual $\hat{\theta}$	$E(\hat{\theta})$	Error estándar $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 y n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ †

* σ_1^2 y σ_2^2 son las varianzas de las poblaciones 1 y 2, respectivamente.

† Se supone que las dos muestras son independientes.

Aun cuando la falta de sesgo es con frecuencia una propiedad deseable para un estimador puntual, no todos los estimadores son insesgados. En el Capítulo 1 definimos la varianza muestral como

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}.$$

Es probable que haya parecido más natural dividir entre n que entre $n - 1$ en la expresión anterior y calcular

$$S'^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n}.$$

El Ejemplo 8.1 establece que S'^2 y S^2 son, respectivamente, estimadores sesgados e insesgados de la varianza poblacional σ^2. Inicialmente identificamos S^2 como la *varianza muestral* porque es un estimador insesgado.

EJEMPLO **8.1** Sea $Y_1, Y_2, \dots, Y_n$ una muestra aleatoria con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$. Demuestre que

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

es un estimador sesgado para σ^2 y que

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

es un estimador insesgado para σ^2.

Solución Se puede demostrar (Ejercicio 1.9) que

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2.$$

En consecuencia,

$$E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right] = E\left(\sum_{i=1}^{n} Y_i^2\right) - nE(\overline{Y}^2) = \sum_{i=1}^{n} E(Y_i^2) - nE(\overline{Y}^2).$$

Observe que $E(Y_i^2)$ es la misma para $i = 1, 2, \dots, n$. Utilicemos esto y el hecho de que la varianza de una variable aleatoria está dada por $V(Y) = E(Y^2) - [E(Y)]^2$ para concluir que $E(Y_i^2) = V(Y_i) + [E(Y_i)]^2 = \sigma^2 + \mu^2$, $E(\overline{Y}^2) = V(\overline{Y}) + [E(\overline{Y})]^2 = \sigma^2/n + \mu^2$, y que

$$\begin{aligned}
E\left[\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right] &= \sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= n\sigma^2 - \sigma^2 = (n-1)\sigma^2.
\end{aligned}$$

Se deduce que

$$E(S'^2) = \frac{1}{n} E\left[\sum_{i=1}^{n} (Y_i - \overline{Y})^2\right] = \frac{1}{n}(n-1)\sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2$$

y que S'^2 está sesgada porque $E(S'^2) \neq \sigma^2$. No obstante,

$$E(S^2) = \frac{1}{n-1} E\left[\sum_{i=1}^{n} (Y_i - \overline{Y})^2\right] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2,$$

por tanto, vemos que S^2 es un estimador insesgado para σ^2. ■

Pueden hacerse dos comentarios finales respecto a los estimadores puntuales de la Tabla 8.1. Primero, los valores esperados y los errores estándar para $\overline{Y}$ y $\overline{Y}_1 - \overline{Y}_2$ dados en la tabla son válidos cualquiera que sea la distribución de la(s) población(es) de donde se tome(n) la(s) muestra(s). En segundo término, los cuatro estimadores poseen distribuciones de probabilidad que son aproximadamente normales para muestras grandes. El teorema del límite central justifica este enunciado para $\overline{Y}$ y $\hat{p}$, y teoremas similares para funciones de medias muestrales justifican la afirmación para $(\overline{Y}_1 - \overline{Y}_2)$ y $(\hat{p}_1 - \hat{p}_2)$. ¿Qué tan grande es "grande"? Para casi todas las poblaciones la distribución de probabilidad de Y tiene forma de campana incluso para muestras relativamente pequeñas (de sólo $n = 5$) y tenderá rápidamente a la normalidad a medida que el tamaño muestral se aproxime a $n = 30$ o más grande. No obstante, a veces es necesario seleccionar muestras más grandes tomadas de poblaciones binomiales porque el tamaño muestral requerido depende de p. La distribución de probabilidad binomial es perfectamente simétrica alrededor de su media cuando $p = 1/2$ y se hace cada vez más asimétrica cuando p tiende a 0 o 1. Como regla general, usted puede suponer que la distribución de $\hat{p}$ tendrá forma de campana y se aproximará a la normalidad para tamaños muestrales tales que $p \pm 3\sqrt{pq/n}$ se encuentre en el intervalo (0, 1), o bien, como ya se demostró en el Ejercicio 7.70, si $n > 9$ (el más grande de p y q)/(el más pequeño de p y q).

Sabemos que Y, $\hat{p}$, $(Y_1 - Y_2)$ y $(\hat{p}_1 - \hat{p}_2)$ son insesgados con distribuciones muestrales casi normales (al menos con forma de campana) para muestras de tamaño moderado; ahora utilicemos esta información para responder algunas preguntas prácticas. Si usamos un estimador una sola vez y obtenemos una sola estimación, ¿qué tan buena será ésta? ¿Cuánto podemos confiar en la validez de nuestra inferencia? Las respuestas a estas preguntas se proporcionan en la sección siguiente.

8.4 Evaluación de la bondad de un estimador puntual

Una forma de evaluar la bondad de cualquier procedimiento de estimación puntual es en términos de las distancias entre las estimaciones que genera y el parámetro objetivo. Esta cantidad, que varía aleatoriamente en muestreo repetido, se denomina *error de estimación.* Por supuesto que nos gustaría que el error de estimación fuera tan pequeño como sea posible.

DEFINICIÓN **8.5**

El *error de estimación* ε es la distancia entre un estimador y su parámetro objetivo. Esto es, $\varepsilon = |\hat{\theta} - \theta|$.

Como $\hat{\theta}$ es una variable aleatoria, el error de estimación también es una cantidad aleatoria y no podemos decir qué tan grande o pequeño será para una estimación particular, pero podemos plantear enunciados de probabilidad acerca de él. Por ejemplo, suponga que $\hat{\theta}$ es un estimador insesgado de θ y tiene una distribución muestral como se muestra en la Figura 8.4. Si seleccionamos dos puntos, $(\theta - b)$ y $(\theta + b)$, situados cerca de las colas de la densidad de probabilidad, la probabilidad de que el error de estimación ε sea menor que b está representada por el área sombreada de la Figura 8.4. Esto es,

$$P(|\hat{\theta} - \theta| < b) = P[-b < (\hat{\theta} - \theta) < b] = P(\theta - b < \hat{\theta} < \theta + b).$$

Podemos considerar a b como un límite probabilístico en el error de estimación. Aun cuando no estamos seguros de que un error determinado sea menor que b, la Figura 8.4 indica que $P(\varepsilon < b)$ es alto. Si b se puede considerar desde un punto de vista práctico como pequeño, entonces $P(\varepsilon < b)$ da una medida de la bondad de una sola estimación. Esta probabilidad identifica la fracción de veces, en muestreo repetido, que el estimador $\hat{\theta}$ cae dentro de b unidades de θ, el parámetro objetivo.

Suponga que deseamos hallar el valor de b para que $P(\varepsilon < b) = .90$. Esto es fácil si conocemos la función de densidad de probabilidad de $\hat{\theta}$. Entonces buscamos un valor b tal que

$$\int_{\theta-b}^{\theta+b} f(\hat{\theta})\, d\hat{\theta} = .90.$$

Pero ya sea que conozcamos o no la distribución de probabilidad de $\hat{\theta}$, si $\hat{\theta}$ es insesgado podemos hallar un límite aproximado en ε al expresar b como un múltiplo del error estándar de $\hat{\theta}$ (recuerde que el error estándar de un estimador es simplemente un nombre alternativo conveniente para la desviación estándar del estimador.) Por ejemplo, para $k \geq 1$, si hacemos $b = k\sigma_{\hat{\theta}}$, sabemos por el teorema de Tchebysheff que ε será menor que $k\sigma_{\hat{\theta}}$ con probabilidad de al menos $1 - 1/k^2$. Un valor de k cómodo y de uso frecuente es $k = 2$. En consecuencia, sabemos que ε será menor que $b = 2\sigma_{\hat{\theta}}$ con probabilidad de al menos .75.

Usted encontrará que, con probabilidad en la cercanía de .95, muchas variables aleatorias observadas en la naturaleza se encuentran a no más de 2 desviaciones estándar de sus medias. La probabilidad de que Y se encuentre en el intervalo $(\mu \pm 2\sigma)$ se muestra en la Tabla 8.2 para

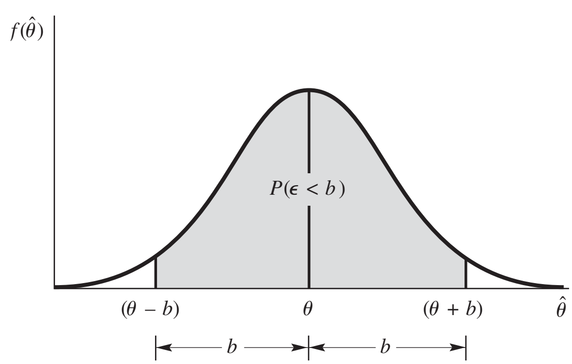


FIGURA **8.4** Distribución muestral de un estimador puntual $\hat{\theta}$

Tabla 8.2 Probabilidad de que $(\mu - 2\sigma) < Y < (\mu + 2\sigma)$

Distribución	Probabilidad
Normal	.9544
Uniforme	1.0000
Exponencial	.9502

las distribuciones de probabilidad normal, uniforme y exponencial. El punto es que $b = 2\sigma_{\hat{\theta}}$ es un buen límite aproximado para el error de estimación en casi todas las situaciones prácticas. De acuerdo con el teorema de Tchebysheff, la probabilidad de que el error de estimación sea menor que este límite es *al menos* .75. Como ya se señaló, los límites para probabilidades dados por el teorema de Tchebysheff suelen ser muy conservadores; las probabilidades reales por lo general exceden los límites de Tchebysheff en una cantidad considerable.

EJEMPLO **8.2** Una muestra de $n = 1000$ votantes, seleccionados al azar en una ciudad, mostró $y = 560$ a favor del candidato Jones. Estime p, la fracción de votantes de la población que están a favor de Jones y precise un límite de error estándar de 2 en el error de estimación.

Solución Utilizaremos el estimador $\hat{p} = Y/n$ para calcular p. Por tanto, la estimación de p, la fracción de votantes que están a favor del candidato Jones, es

$$\hat{p} = \frac{y}{n} = \frac{560}{1000} = .56.$$

¿Qué tan confiable es este valor? La distribución de probabilidad de $\hat{p}$ se aproxima en forma muy precisa mediante una distribución de probabilidad normal para muestras grandes. Como $n = 1000$, cuando $b = 2\sigma_{\hat{p}}$, la probabilidad de que ε sea menor que b es aproximadamente .95.

De la Tabla 8.1, el error estándar del estimador para p está dado por $\sigma_{\hat{p}} = \sqrt{pq/n}$. Por tanto,

$$b = 2\sigma_{\hat{p}} = 2\sqrt{\frac{pq}{n}}.$$

Desafortunadamente, para calcular b necesitamos conocer p y determinar el valor de p fue el objetivo de nuestro muestreo. No obstante, este empate aparente no es una desventaja porque $\sigma_{\hat{p}}$ varía poco para cambios pequeños de p. Por tanto, la sustitución de la estimación $\hat{p}$ por p produce un error pequeño en el cálculo del valor exacto de $b = 2\sigma_{\hat{p}}$. Entonces, para nuestro ejemplo, tenemos

$$b = 2\sigma_{\hat{p}} = 2\sqrt{\frac{pq}{n}} \approx 2\sqrt{\frac{(.56)(.44)}{1000}} = .03.$$

¿Cuál es la importancia de nuestros cálculos? La probabilidad de que el error de estimación sea menor que .03 es aproximadamente .95. En consecuencia, podemos tener una confianza razonable de que nuestra estimación, .56, está a no más de .03 del valor verdadero de p, la proporción de votantes en la población que está a favor de Jones. ■

EJEMPLO **8.3** Una comparación de la durabilidad de dos tipos de llantas para automóvil se obtuvo de muestras de pruebas en carretera de $n_1 = n_2 = 100$ llantas de cada tipo. Se registró el número de millas hasta quedar inútiles, el desgaste se definió como el número de millas hasta que la cantidad restante de superficie de rodamiento llegó a un valor pequeño especificado previamente. Las mediciones para los dos tipos de llantas se obtuvieron de manera independiente y se calcularon las siguientes medias y varianzas:

$$\overline{y}_1 = 26{,}400 \text{ millas}, \qquad \overline{y}_2 = 25{,}100 \text{ millas},$$
$$s_1^2 = 1{,}440{,}000, \qquad s_2^2 = 1{,}960{,}000.$$

Estime la diferencia en la media de millas hasta quedar inútiles y precise un límite de error estándar de 2 en el error de estimación.

Solución La estimación puntual de $(\mu_1 - \mu_2)$ es

$$(\overline{y}_1 - \overline{y}_2) = 26{,}400 - 25{,}100 = 1300 \text{ millas},$$

y el error estándar del estimador (vea Tabla 8.1) es

$$\sigma_{(\overline{Y}_1 - \overline{Y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Debemos conocer σ_1^2 y σ_2^2, o tener buenos valores aproximados de ellas para calcular $\sigma_{(\overline{Y}_1 - \overline{Y}_2)}$. Con frecuencia se pueden calcular valores razonablemente precisos de σ_1^2 y σ_2^2 a partir de datos experimentales similares, recolectados algún tiempo antes, o se pueden obtener de datos muestrales actuales mediante el uso de los estimadores insesgados

$$\hat{\sigma}_i^2 = S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2, \qquad i = 1, 2.$$

Estas estimaciones serán adecuadas si los tamaños muestrales son razonablemente grandes, por ejemplo $n_i \geq 30$, para $i = 1, 2$. Los valores calculados de S_1^2 y S_2^2, basados en las dos pruebas de desgaste, son $s_1^2 = 1{,}440{,}000$ y $s_2^2 = 1{,}960{,}000$. Sustituyendo estos valores por σ_1^2 y σ_2^2 en la fórmula para $\sigma_{(\overline{Y}_1 - \overline{Y}_2)}$, tenemos

$$\sigma_{(\overline{Y}_1 - \overline{Y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1{,}440{,}000}{100} + \frac{1{,}960{,}000}{100}}$$
$$= \sqrt{34{,}000} = 184.4 \text{ millas}.$$

En consecuencia, estimamos que la diferencia en desgaste medio es de 1300 millas y esperamos que el error de estimación sea menor que $2\sigma_{(\overline{Y}_1 - \overline{Y}_2)}$, o sea 368.8 millas, con una probabilidad de aproximadamente .95. ■

Ejercicios

8.21 Un investigador está interesado en la posibilidad de unir las aptitudes de televisión e Internet. Una muestra aleatoria de $n = 50$ usuarios de Internet dio que el tiempo medio semanal empleado en ver televisión era de 11.5 horas y que la desviación estándar era de 3.5 horas. Estime el tiempo medio poblacional que los usuarios de Internet pasan viendo televisión y fije un límite para el error de estimación.

8.22 Es frecuente que un aumento en el porcentaje de ahorros de los consumidores se encuentre ligado a la falta de confianza en la economía y se dice que es un indicador de una tendencia recesiva en aquélla. Un muestreo aleatorio de $n = 200$ cuentas de ahorros en una comunidad local mostró que el aumento medio en los valores de las cuentas de ahorros era de 7.2% en los últimos 12 meses, con desviación estándar de 5.6%. Estime el porcentaje medio de aumento en los valores de las cuentas de ahorros en los últimos 12 meses para depositantes de la comunidad. Establezca un límite para su error de estimación.

8.23 La Environmental Protection Agency en conjunto con la Universidad de Florida recientemente realizó un amplio estudio de los posibles efectos de trazas de elementos en el agua potable sobre la formación de cálculos renales. La tabla siguiente presenta datos de edad, cantidad de calcio en agua potable (medida en partes por millón) y hábitos de fumar. Estos datos se obtuvieron de individuos con problemas actuales de cálculos renales, todos los cuales vivían en las dos Carolinas y en estados de las Montañas Rocallosas.

	Carolinas	Rocallosas
Tamaño muestral	467	191
Edad promedio	45.1	46.4
Desviación estándar de edad	10.2	9.8
Componente medio de calcio (ppm)	11.3	40.1
Desviación estándar del calcio	16.6	28.4
Proporción de fumadores en el momento del estudio	.78	.61

a Estime la concentración promedio de calcio en el agua potable para pacientes con cálculos renales en las Carolinas. Establezca un límite para el error de estimación.

b Calcule la diferencia en edades medias para pacientes con cálculos renales en las Carolinas y en las Rocallosas. Fije un límite para el error de estimación.

c Calcule y precise un límite de desviación estándar de 2 en la diferencia en proporciones de pacientes con cálculos renales, de las Carolinas y las Rocallosas, que eran fumadores en el momento de hacer el estudio.

8.24 Los resultados de una encuesta de opinión pública publicados en la Internet[2] indicaron que 69% de quienes respondieron clasificaron el costo de la gasolina como una crisis o problema importante. El artículo indica que 1001 adultos, de 18 años o más, fueron entrevistados y que los resultados tienen un error de muestreo de 3%. ¿Cómo se calculó el 3% y cómo debería interpretarse? ¿Podemos concluir que una mayoría de los individuos del grupo de mayores de 18 años pensaron que el costo de la gasolina era una crisis o problema importante?

8.25 Se realizó un estudio para comparar el promedio de llamadas de emergencia a la policía en cada turno de 8 horas en dos distritos de una ciudad grande. Se seleccionaron aleatoriamente muestras de 100 turnos de 8 horas de los registros policíacos para cada una de las dos regiones y se registró el número de llamadas de emergencia para cada turno. Los estadísticos muestrales se proporcionan en la tabla siguiente.

	Región	
	1	2
Tamaño muestral	100	100
Media muestral	2.4	3.1
Varianza muestral	1.44	2.64

2 *Fuente:* Jeffrey M. Jones, “CNN/USA Today/Gallup Poll—Public Expects Gas Price Increase To Be Permanent,” http:www.gallup.com/content/print/.aspx?ci=11257,8 de abril de 2004.

a Calcule la diferencia en el número medio de llamadas de emergencia a la policía por turno de 8 horas entre los dos distritos de la ciudad.

b Encuentre un límite para el error de estimación.

8.26 Los vehículos gemelos *Spirit* y *Opportunity*, que recorrieron la superficie de Marte en el invierno de 2004, hallaron evidencia de que una vez hubo agua en el planeta, elevando la posibilidad de que una vez hubo vida en el mismo. ¿Piensa usted que Estados Unidos debería continuar un programa para enviar seres humanos a Marte? Una encuesta de opiniones[3] indicó que 49% de los 1093 adultos encuestados piensa que se debería continuar ese programa.

a Estime la proporción de todos los norteamericanos que piensan que Estados Unidos debería continuar un programa para enviar seres humanos a Marte. Encuentre un límite en el error de estimación.

b La encuesta en realidad formuló varias preguntas. Si quisiéramos informar un error de estimación que sería válido para todas las preguntas de la encuesta, ¿qué valor deberíamos usar? [*Sugerencia:* ¿cuál es el máximo valor posible para $p \times q$?]

8.27 Una muestra aleatoria de 985 "probables votantes" —a quienes se considera que probablemente voten en una elección próxima— fueron encuestados durante una indagación por teléfono realizada por el partido republicano. De aquellos contactados, 592 indicaron que tenían intención de votar por el candidato republicano en la elección.

a De acuerdo con este estudio, la estimación para p, la proporción de todos los "posibles votantes" que votarán por el candidato republicano, es $p = .601$. Encuentre un límite para el error de estimación.

b Si los "probables votantes" son representativos de quienes realmente votarán, ¿piensa usted que el candidato republicano será elegido? ¿Por qué? ¿Qué confianza tiene usted en su decisión?

c ¿Puede usted dar razones por las que los encuestados podrían no ser representativos de los que en realidad voten en la elección?

8.28 En un estudio de la relación entre orden de nacimiento y éxito universitario, un investigador encontró que 126 en una muestra de 180 egresados de universidad fueron hijos primogénitos o hijos únicos; en una muestra de 100 no egresados de edad y antecedentes socio económicos comparables, el número de primogénitos o hijos únicos era de sólo 54. Estime la diferencia en las proporciones de primogénitos o hijos únicos para las dos poblaciones de las que se tomaron estas muestras. Establezca un límite para el error de estimación.

8.29 En ocasiones las encuestas dan información interesante acerca de cuestiones que no parecieron ser de interés en la encuesta inicialmente. Los resultados de dos encuestas de la CNN/USA Today/Gallup, una de ellas efectuada en marzo de 2003 y otra en noviembre de 2003, fueron presentados recientemente en línea.[4] Ambas encuestas comprendieron muestras de 1001 adultos, de 18 años de edad o más. En la muestra de marzo, 45% de los muestreados dijeron ser aficionados al beisbol profesional mientras que 51% de los encuestados en noviembre dijeron ser aficionados.

a Dé una estimación puntual para la diferencia en las proporciones de estadounidenses que dicen ser aficionados al beisbol en marzo (al principio de la temporada) y noviembre (después de la serie mundial). Proporcione un límite para el error de estimación.

b ¿Hay suficiente evidencia para concluir que el apoyo de los aficionados es mayor al final de la temporada? Explique.

3 *Fuente:* "Space Exploration," Associated Press Poll, http:www.pollingreport.com/science.htm#Space, 5 de abril de 2004.

4 *Fuente*: Mark Gillespie, "Baseball Fans Overwhelmingly Want Mandatory Steroid Testing," http:www.gallup.com/content/print.aspx?ci=11245, 14 de febrero de 2004.

8.30 Consulte el Ejercicio 8.29. Dé la estimación puntual y un límite para el error de estimación de la proporción de adultos que hubieran dicho que eran aficionados al beisbol en marzo de 2003. ¿Es probable que el valor de su estimación se desvíe hasta en 10%? ¿Por qué?

8.31 En un estudio para comparar los efectos percibidos de dos calmantes para el dolor, a 200 adultos seleccionados aleatoriamente se les dio el primer calmante y 93% indicaron un alivio considerable del dolor. De los 450 a quienes se dio otro calmante para el dolor, 96% indicaron haber experimentado mejoría apreciable.

a Dé una estimación para la diferencia en las proporciones de todos los adultos que indicarían sentir alivio del dolor percibido después de tomar los dos calmantes. Precise un límite para el error de estimación.

b Con base en su respuesta al inciso a, ¿existe evidencia de que las proporciones de los que experimentan alivio difieran de los que toman los dos calmantes para el dolor? ¿Por qué?

8.32 Un auditor muestrea aleatoriamente 20 cuentas por cobrar de entre 500 de esas cuentas de la empresa de un cliente. El auditor hace una lista de la cantidad de cada cuenta y verifica si los documentos que sirven de base cumplen con los procedimientos estipulados. Los datos se registran en la tabla siguiente (las cantidades son en dólares, Y = sí y N = no).

Cuenta	Cantidad	Cumple	Cuenta	Cantidad	Cumple
1	278	Y	11	188	N
2	192	Y	12	212	N
3	310	Y	13	92	Y
4	94	N	14	56	Y
5	86	Y	15	142	Y
6	335	Y	16	37	Y
7	310	N	17	186	N
8	290	Y	18	221	Y
9	221	Y	19	219	N
10	168	Y	20	305	Y

Estime el total de cuentas por cobrar para las 500 cuentas de la empresa y fije un límite para el error de estimación. ¿Piensa usted que el *promedio* de cuentas por cobrar para la empresa es mayor a $250 dólares? ¿Por qué?

8.33 Consulte el Ejercicio 8.32. De los datos obtenidos en las revisiones de cumplimiento, estime la proporción de las cuentas de la empresa que no cumplen con los procedimientos estipulados. Precise un límite para el error de estimación. ¿Piensa usted que la proporción de cuentas que cumplen con los procedimientos estipulados es mayor que 80%? ¿Por qué?

8.34 Podemos definir un límite de desviación estándar 2 en el error de estimación con cualquier estimador para el cual podamos hallar una estimación razonable del error estándar. Suponga que $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria de una distribución de Poisson con media λ. Sabemos que $V(Y_i) = \lambda$, y por tanto que $E(\overline{Y}) = \lambda$ y $V(\overline{Y}) = \lambda/n$. ¿Cómo emplearía usted $Y_1, Y_2, \ldots, Y_n$ para estimar λ? ¿Cómo estimaría el error estándar de su estimador?

8.35 Consulte el Ejercicio 8.34. En el aluminio policristalino, la cantidad de puntos de nucleación de grano por unidad de volumen está modelada con una distribución de Poisson con media λ. Cincuenta especímenes de prueba de volumen unitario sometidos a recocido con el régimen A produjeron un promedio de 20 puntos por volumen unitario. Cincuenta especímenes de prueba de volumen unitario seleccionados de manera independiente sometidos a recocido con el régimen B produjeron un promedio de 23 puntos por volumen unitario.

a Estime la media λ_A del número de puntos de nucleación para el régimen A y precise un límite de error estándar de 2 en el error de estimación.

b Estime la diferencia en la media del número de puntos de nucleación $\lambda_A - \lambda_B$ para los regímenes A y B. Establezca un límite de error estándar de 2 para el error de estimación. ¿Se diría que el régimen B tiende a producir un promedio mayor de puntos de nucleación? ¿Por qué?

8.36 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución exponencial con media θ, entonces $E(Y_i) = \theta$ y $V(Y_i) = \theta^2$. Entonces, $E(\overline{Y}) = \theta$ y $V(\overline{Y}) = \theta^2/n$ o $\sigma_{\overline{Y}} = \theta/\sqrt{n}$. Sugiera un estimador insesgado para θ y dé una estimación para el error estándar del estimador sugerido.

8.37 Consulte el Ejercicio 8.36. Un ingeniero observa $n = 10$ mediciones independientes de la duración de un componente electrónico. El promedio de estas 10 mediciones es de 1020 horas. Si estas duraciones provienen de una distribución exponencial con media θ, estime θ y ponga un límite de error estándar de 2 en el error de estimación.

8.38 El número de personas que acuden a un banco de sangre hasta que se encuentra la primera de ellas con sangre tipo A es una variable aleatoria Y con distribución geométrica. Si p denota la probabilidad de que cualquier persona seleccionada aleatoriamente posea sangre tipo A, entonces $E(Y) = 1/p$ y $V(Y) = (1 - p)/p^2$.

a Determine una función de Y que sea un estimador insesgado de $V(Y)$.

b Sugiera una forma de establecer un límite de error estándar de 2 para el error de estimación cuando Y se use para calcular $1/p$.

8.5 Intervalos de confianza

Un *estimador de intervalo* es una regla que especifica el método para usar las mediciones muestrales en el cálculo de dos números que forman los puntos extremos del intervalo. En el caso ideal, el intervalo resultante tiene dos propiedades: primero, contiene el parámetro objetivo θ; en segundo, su amplitud será relativamente pequeña. Uno o ambos puntos extremos del intervalo, siendo funciones de las mediciones muéstrales, variarán aleatoriamente de una muestra a otra. Entonces, la longitud y ubicación del intervalo son cantidades aleatorias; no podemos estar seguros de que el parámetro objetivo θ (fijo) caiga entre los puntos extremos de cualquier intervalo individual calculado a partir de una sola muestra. En este caso, nuestro objetivo es hallar un estimador de intervalo capaz de generar intervalos estrechos que tengan una alta probabilidad de incluir a θ.

Los estimadores de intervalo suelen recibir el nombre de *intervalos de confianza*. Los puntos extremos superior e inferior de un intervalo de confianza se denominan *límites de confianza superior* e *inferior*, respectivamente. La probabilidad de que un intervalo de confianza (aleatorio) incluya a θ (una cantidad fija) se llama *coeficiente de confianza.* Desde un punto de vista práctico, el coeficiente de confianza identifica la fracción de veces, en muestreo repetido, que los intervalos construidos contienen al parámetro objetivo θ. Si sabemos que el coeficiente de confianza asociado con nuestro estimador es alto, podemos estar suficentemente seguros de que cualquier intervalo de confianza, construido con el uso de los resultados de una sola muestra, contendrá a θ.

Suponga que $\hat{\theta}_L$ y $\hat{\theta}_U$ son los límites de confianza (aleatorios) superior e inferior, respectivamente, para un parámetro θ. Entonces, si

$$P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U\right) = 1 - \alpha,$$

la probabilidad $(1 - \alpha)$ es el *coeficiente de confianza*. El intervalo aleatorio resultante definido por $\left[\hat{\theta}_L, \hat{\theta}_U\right]$ se denomina *intervalo de confianza bilateral.*

También es posible formar un *intervalo de confianza unilateral* tal que

$$P\left(\hat{\theta}_L \leq \theta\right) = 1 - \alpha.$$

Aun cuando sólo $\hat{\theta}_L$ es aleatorio en este caso, el intervalo de confianza es $[\hat{\theta}_L, \infty)$. Del mismo modo, podríamos tener un intervalo de confianza unilateral superior tal que

$$P(\theta \leq \hat{\theta}_U) = 1 - \alpha.$$

El intervalo de confianza implicado aquí es $(-\infty, \hat{\theta}_U]$.

Un método muy útil para encontrar intervalos de confianza se llama *método del pivote*. Éste consiste en determinar una cantidad que actúe como pivote y que posea las dos características siguientes:

1. Que sea una función de las medidas muestrales y el parámetro desconocido θ, donde θ sea la *única* cantidad desconocida.
2. Que su distribución de probabilidad no dependa del parámetro θ.

Si se conoce la distribución de probabilidad de la cantidad que actúa como pivote, el siguiente procedimiento lógico puede usarse para obtener la estimación por intervalos deseada. Si Y es cualquier variable aleatoria, $c > 0$ es una constante y $P(a \leq Y \leq b) = .7$; entonces ciertamente $P(ca \leq cY \leq cb) = .7$. Del mismo modo, para cualquier constante d, $P(a + d \leq Y + d \leq b + d) = .7$. Esto es, la probabilidad del evento $(a \leq Y \leq b)$ no resulta afectada por un cambio de escala o una traslación de Y. Entonces, si conocemos la distribución de probabilidad de una cantidad pivote, podemos usar operaciones como éstas para formar la estimación por intervalos que buscamos. Ilustramos este método con los ejemplos siguientes.

EJEMPLO **8.4** Suponga que obtenemos una sola observación Y de una distribución exponencial con media θ. Use Y para construir un intervalo de confianza para θ con un coeficiente de confianza de .90.

Solución La función de densidad de probabilidad para Y está dada por

$$f(y) = \begin{cases} \left(\dfrac{1}{\theta}\right) e^{-y/\theta}, & y \geq 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por el método de transformaciones del Capítulo 6 podemos ver que $U = Y/\theta$ tiene la función de densidad exponencial dada por

$$f_U(u) = \begin{cases} e^{-u}, & u > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

La función de densidad para U aparece graficada en la Figura 8.5. $U = Y/\theta$ es una función de Y (la medición muestral) y θ, y la distribución de U no depende de θ. Entonces, podemos emplear $U = Y/\theta$ como cantidad pivote. Como buscamos un estimador de intervalo con coeficiente de confianza igual a .90, encontramos dos números a y b tales que

$$P(a \leq U \leq b) = .90.$$

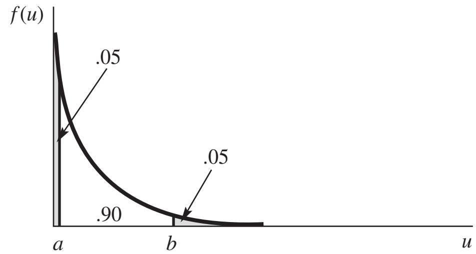


FIGURA **8.5** Función de densidad para U, Ejemplo 8.4

Una forma de hacer esto es elegir a y b para satisfacer

$$P(U < a) = \int_0^a e^{-u}\,du = .05 \quad \text{y} \quad P(U > b) = \int_b^\infty e^{-u}du = .05.$$

Estas ecuaciones dan como resultado

$$1 - e^{-a} = .05 \ \text{ y } \ e^{-b} = .05 \quad \text{o bien, } a = .051, \quad b = 2.996.$$

Por consiguiente

$$.90 = P(.051 \leq U \leq 2.996) = P\left(.051 \leq \frac{Y}{\theta} \leq 2.996\right).$$

Como estamos buscando un estimador de intervalo para θ, manipulamos las desigualdades que describen el evento para aislar θ en el centro. Y tiene una distribución exponencial, de modo que $P(Y > 0) = 1$ y mantenemos la dirección de las desigualdades si dividimos todo entre Y. Esto es,

$$.90 = P\left(.051 \leq \frac{Y}{\theta} \leq 2.996\right) = P\left(\frac{.051}{Y} \leq \frac{1}{\theta} \leq \frac{2.996}{Y}\right).$$

Tomando recíprocos (y por tanto invirtiendo la dirección de las desigualdades) obtenemos

$$.90 = P\left(\frac{Y}{.051} \geq \theta \geq \frac{Y}{2.996}\right) = P\left(\frac{Y}{2.996} \leq \theta \leq \frac{Y}{.051}\right).$$

Entonces, vemos que $Y/2.996$ y $Y/.051$ forman los límites de confianza inferior y superior, respectivamente, que estábamos buscando. Para obtener los valores numéricos de estos límites debemos observar un valor real para Y y sustituirlo en las fórmulas dadas para los límites de confianza. Sabemos que límites de la forma $(Y/2.996, Y/.051)$ incluirán los valores (desconocidos) verdaderos de θ para 90% de los valores de Y que obtendríamos por muestreo repetido a partir de esta distribución exponencial. ■

EJEMPLO **8.5** Suponga que tomamos una muestra de tamaño $n = 1$ de una distribución uniforme definida en el intervalo $[0, \theta\,]$, donde θ es desconocida. Encuentre un límite de confianza inferior de 95% para θ.

Solución Como Y es uniforme en $[0, \theta]$, los métodos del Capítulo 6 se pueden usar para demostrar que $U = Y/\theta$ está uniformemente distribuida en $[0, 1]$. Esto es,

$$f_U(u) = \begin{cases} 1, & 0 \leq u \leq 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

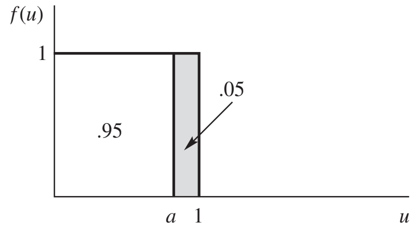


FIGURA **8.6** Función de densidad para U, Ejemplo 8.5

La figura 8.6 contiene una gráfica de la función de densidad para U. De nuevo, vemos que U satisface los requisitos de una cantidad pivote. Como buscamos un límite de confianza inferior de 95% para θ, determinamos el valor para a de modo que $P(U \leq a) = .95$. Esto es,

$$\int_0^a (1)\, du = .95,$$

o sea que $a = .95$. Por tanto,

$$P(U \leq .95) = P\left(\frac{Y}{\theta} \leq .95\right) = P(Y \leq .95\theta) = P\left(\frac{Y}{.95} \leq \theta\right) = .95.$$

Vemos que $Y/.95$ es un límite de confianza inferior para θ, con coeficiente de confianza .95. Como cualquier Y observada debe ser menor que θ, es intuitivamente razonable tener el límite de confianza inferior para θ ligeramente mayor que el valor observado de Y. ■

Los dos ejemplos anteriores ilustran el uso del método del pivote para determinar límites de confianza para parámetros desconocidos. En cada caso las estimaciones de intervalo se desarrollaron con base en una sola observación de la distribución. Estos ejemplos se presentaron básicamente para ilustrar el método del pivote. En las secciones restantes de este capítulo usaremos este método en coordinación con las distribuciones de muestreo presentadas en el Capítulo 7 para desarrollar algunas estimaciones por intervalos de mayor importancia práctica.

Ejercicios

8.39 Suponga que la variable aleatoria Y tiene una distribución gamma con parámetros $\alpha = 2$ y β desconocida. En el Ejercicio 6.46 usted utilizó el método de las funciones generadoras de momento para demostrar un resultado general que implicaba que $2Y/\beta$ tiene una distribución χ^2 con 4 grados de libertad. Usando $2Y/\beta$ como cantidad pivote, deduzca un intervalo de confianza de 90% para β.

8.40 Suponga que la variable aleatoria Y es una observación de una distribución normal con media μ desconocida y varianza 1. Encuentre un

a intervalo de confianza de 95% para μ,

b límite de confianza superior de 95% para μ,

c límite de confianza inferior de 95% para μ.

8.41 Suponga que Y está distribuida normalmente con media 0 y varianza σ^2 desconocida. Entonces Y^2/σ^2 tiene una distribución χ^2 con grado de libertad 1. Use la cantidad pivote Y^2/σ^2 para hallar un

a intervalo de confianza de 95% para σ^2,
b límite de confianza superior de 95% para σ^2,
c límite de confianza inferior de 95% para σ^2.

8.42 Use las respuestas del Ejercicio 8.41 para hallar un

a intervalo de confianza de 95% para σ,
b límite de confianza superior de 95% para σ,
c límite de confianza inferior de 95% para σ.

8.43 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población con distribución uniforme en el intervalo $(0, \theta)$. Sean $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ y $U = (1/\theta)Y_{(n)}$.

a Demuestre que U tiene función de distribución

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u^n, & 0 \le u \le 1, \\ 1, & u > 1. \end{cases}$$

b Como la distribución de U no depende de θ, U es una cantidad pivote. Encuentre un límite de confianza inferior de 95% para θ.

8.44 Suponga que Y tiene la siguiente función de densidad de probabilidad

$$f_Y(y) = \begin{cases} \dfrac{2(\theta - y)}{\theta^2}, & 0 < y < \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que Y tiene función de distribución

$$F_Y(y) = \begin{cases} 0, & y \le 0, \\ \dfrac{2y}{\theta} - \dfrac{y^2}{\theta^2}, & 0 < y < \theta, \\ 1, & y \ge \theta. \end{cases}$$

b Demuestre que Y/θ es una cantidad pivote.
c Use la cantidad pivote del inciso b para hallar un límite de confianza inferior de 90% para θ.

8.45 Consulte el Ejercicio 8.44.

a Use la cantidad pivote del Ejercicio 8.44(b) para hallar un límite de confianza superior de 90% para θ.
b Si $\hat{\theta}_L$ es el límite de confianza inferior para θ obtenido en el Ejercicio 8.44(c) y $\hat{\theta}_U$ es el límite superior hallado en el inciso a, ¿cuál es el coeficiente de confianza del intervalo $(\hat{\theta}_L,\ \hat{\theta}_U)$?

8.46 Consulte el Ejemplo 8.4 y suponga que Y es una sola observación de una distribución exponencial con media θ.

a Use el método de las funciones generadoras de momento para demostrar que $2Y/\theta$ es una cantidad pivote y tiene una distribución χ^2 con 2 grados de libertad.
b Use la cantidad pivote $2Y/\theta$ para deducir un intervalo de confianza de 90% para θ.
c Compare el intervalo que obtuvo en el inciso b con el intervalo obtenido en el Ejemplo 8.4.

8.47 Consulte el Ejercicio 8.46. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra de tamaño n para una distribución exponencial con media θ.

a Utilice el método de las funciones generadoras de momento para demostrar que $2\sum_{i=1}^{n} Y_i/\theta$ es una cantidad pivote y tiene una distribución χ^2 con $2n$ grados de libertad.
b Use la cantidad pivote $2\sum_{i=1}^{n} Y_i/\theta$ para deducir un intervalo de confianza de 95% para θ.

c Si una muestra de tamaño $n = 7$ da $\overline{y} = 4.77$, use el resultado del inciso b para dar un intervalo de confianza de 95% para θ.

8.48 Consulte los Ejercicios 8.39 y 8.47. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra de tamaño n de una población con distribución gamma con $\alpha = 2$ y β desconocida.

a Utilice el método de las funciones generadoras de momento para demostrar que $2\sum_1^n Y_i/\beta$ es una cantidad pivote y tiene una distribución χ^2 con $4n$ grados de libertad.

b Use la cantidad pivote $2\sum_1^n Y_i/\beta$ para deducir un intervalo de confianza de 95% para β.

c Si una muestra de tamaño $n = 5$ da $\overline{y} = 5.39$, use el resultado del inciso b para dar un intervalo de confianza para β.

8.49 Consulte el Ejercicio 8.48. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra de tamaño n de una población con distribución gamma con parámetros α y β.

a Si $\alpha = m$, donde m es un entero conocido y β es una incógnita, encuentre una cantidad pivote que tenga una distribución χ^2 con $m \times n$ grados de libertad. Use esta cantidad pivote para deducir un intervalo de confianza de $100(1 - \alpha)\%$ para β.

b Si $\alpha = c$, donde c es una constante conocida pero no un entero y β es una incógnita, encuentre una cantidad pivote que tenga una distribución gamma con parámetros $\alpha^\star = cn$ y $\beta^\star = 1$. Dé una fórmula para un intervalo de confianza $100(1 - \alpha)\%$ para β.

c **Ejercicio Applet** Consulte el inciso b. Si $\alpha = c = 2.57$ y una muestra de tamaño $n = 10$ da $\overline{y} = 11.36$, dé un intervalo de confianza de 95% para β. [Use la aplicación *Gamma Probabilities and Quantiles* para obtener cuantiles apropiados para la cantidad pivote que obtuvo en el inciso b.]

8.6 Intervalos de confianza en una muestra grande

En la Sección 8.3 presentamos algunos estimadores puntuales insesgados para los parámetros μ, p, $\mu_1 - \mu_2$, y $p_1 - p_2$. Como indicamos en esa sección, para muestras grandes todos estos estimadores puntuales tienen distribuciones muestrales aproximadamente normales con errores estándar como los que se dan en la Tabla 8.1. Esto es, dadas las condiciones de la Sección 8.3, si el parámetro objetivo θ es μ, p, $\mu_1 - \mu_2$, o $p_1 - p_2$, entonces para muestras grandes,

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

posee aproximadamente una distribución normal estándar. En consecuencia, $Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$ forma (al menos aproximadamente) una cantidad pivote y el método del pivote se puede emplear para desarrollar intervalos de confianza para el parámetro objetivo θ.

EJEMPLO **8.6** Sea $\hat{\theta}$ un estadístico que está normalmente distribuido con media θ y error estándar $\sigma_{\hat{\theta}}$. Encuentre un intervalo de confianza para θ que posea un coeficiente de confianza igual a $(1 - \alpha)$.

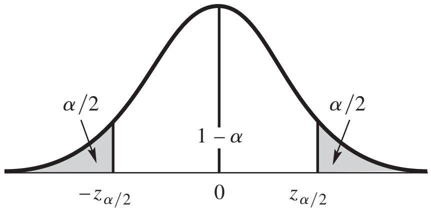
Solución La cantidad

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

tiene una distribución normal estándar. Ahora seleccione dos valores de las colas de esta distribución, $z_{\alpha/2}$ y $-z_{\alpha/2}$, tales que (vea Figura 8.7)

$$P(-z_{\alpha/2} \le Z \le z_{/\alpha 2}) = 1 - \alpha.$$

FIGURA **8.7**
Ubicación de $z_{\alpha/2}$ y $-z_{\alpha/2}$



Sustituyendo por Z en el enunciado de probabilidad, tenemos

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) = 1-\alpha.$$

Multiplicando por $\sigma_{\hat{\theta}}$, obtenemos

$$P(-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \hat{\theta}-\theta \leq z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$$

y restando $\hat{\theta}$ de cada término de la desigualdad, obtenemos

$$P(-\hat{\theta}-z_{\alpha/2}\sigma_{\hat{\theta}} \leq -\theta \leq -\hat{\theta}+z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha.$$

Por último, multiplicando cada término por –1 y, en consecuencia, cambiando la dirección de las desigualdades, tenemos

$$P(\hat{\theta}-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta}+z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha.$$

Entonces, los puntos extremos para un intervalo de confianza de $100(1-\alpha)\%$ para θ están dados por

$$\hat{\theta}_L = \hat{\theta}-z_{\alpha/2}\sigma_{\hat{\theta}} \quad \text{y} \quad \hat{\theta}_U = \hat{\theta}+z_{\alpha/2}\sigma_{\hat{\theta}}.$$

■

Por medio de argumentos análogos podemos determinar que límites de confianza unilaterales de $100(1-\alpha)\%$, a menudo llamados límites superior e inferior, respectivamente, están dados por

límite inferior al $100(1-\alpha)\%$ para $\theta = \hat{\theta} - z_{\alpha}\sigma_{\hat{\theta}}$,
límite superior al $100(1-\alpha)\%$ para $\theta = \hat{\theta} + z_{\alpha}\sigma_{\hat{\theta}}$,

Suponga que calculamos un límite inferior al $100(1-\alpha)\%$ y un límite superior al $100(1-\alpha)\%$ para θ. Entonces decidimos usarlos para formar un intervalo de confianza para θ. ¿Cuál será el coeficiente de confianza de este intervalo? Una rápida mirada a lo anterior confirma que combinar límites inferior y superior, cada uno con coeficiente de confianza $1-\alpha$, genera un intervalo bilateral con coeficiente de confianza de $1-2\alpha$.

De acuerdo con las condiciones descritas en la Sección 8.3, los resultados obtenidos en esta sección se pueden usar para hallar intervalos de confianza de una muestra grande (unilateral o bilateral) para μ, p, $(\mu_1-\mu_2)$, y (p_1-p_2). Los siguientes ejemplos ilustran aplicaciones del método general desarrollado en el Ejemplo 8.6.

EJEMPLO **8.7** Se registraron los tiempos de compra de $n = 64$ clientes seleccionados al azar en un supermercado local. El promedio y varianza de los 64 tiempos de compra fueron 33 minutos y 256 minutos^2, respectivamente. Estime μ, el verdadero promedio de tiempo de compra por cliente, con un coeficiente de confianza de $1-\alpha = .90$.

Solución En este caso estamos interesados en el parámetro $\theta = \mu$. Entonces, $\hat{\theta} = \overline{y} = 33$ y $s^2 = 256$ para una muestra de $n = 64$ tiempos de compra. La varianza poblacional σ^2 es desconocida, de modo que (como en la Sección 8.3) usamos s^2 como su valor estimado. El intervalo de confianza

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

tiene la forma

$$\overline{y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \overline{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right).$$

De la Tabla 4, Apéndice 3, $z_{\alpha/2} = z_{.05} = 1.645$; por tanto, los límites de confianza están dados por

$$\overline{y} - z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 33 - 1.645\left(\frac{16}{8}\right) = 29.71,$$

$$\overline{y} + z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 33 + 1.645\left(\frac{16}{8}\right) = 36.29.$$

Así, nuestro intervalo de confianza para μ es (29.71, 36.29). En muestreo repetido, aproximadamente 90% de todos los intervalos de la forma $\overline{Y} \pm 1.645(S/\sqrt{n})$ incluyen μ, el tiempo medio real de compra por cliente. Aun cuando no sabemos si el intervalo particular (29.71, 36.29) contiene a μ, el procedimiento que lo generó proporciona intervalos que captan la media real en aproximadamente 95% de todos los ejemplos en los que se utilice el procedimiento. ■

EJEMPLO **8.8** Dos marcas de refrigeradores, denotadas por A y B, están garantizadas por 1 año. En una muestra aleatoria de 50 refrigeradores de la marca A, se observó que 12 de ellos fallaron antes de terminar el periodo de garantía. Una muestra aleatoria independiente de 60 refrigeradores de la marca B también reveló 12 fallas durante el período de garantía. Calcule la diferencia real $(p_1 - p_2)$ entre las proporciones de fallas durante el período de garantía, con un coeficiente de confianza de aproximadamente .98.

Solución El intervalo de confianza

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

tiene ahora la forma

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}.$$

Debido a que p_1, q_1, p_2 y q_2 son desconocidos, el valor exacto de $\sigma_{\hat{\theta}}$ no se puede evaluar. Pero, como se indica en la Sección 8.3, podemos obtener una buena aproximación para $\sigma_{\hat{\theta}}$ al sustituir $\hat{p}_1$, $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{p}_2$ y $\hat{q}_2 = 1 - \hat{p}_2$ por p_1, q_1, p_2 y q_2, respectivamente.

Para este ejemplo, $\hat{p}_1 = .24$, $\hat{q}_1 = .76$, $\hat{p}_2 = .20$, $\hat{q}_2 = .80$ y $z_{.01} = 2.33$. El intervalo de confianza de 98% deseado es

$$(.24 - .20) \pm 2.33\sqrt{\frac{(.24)(.76)}{50} + \frac{(.20)(.80)}{60}}$$

$$.04 \pm .1851 \quad \text{o} \quad [-.1451,\ .2251].$$

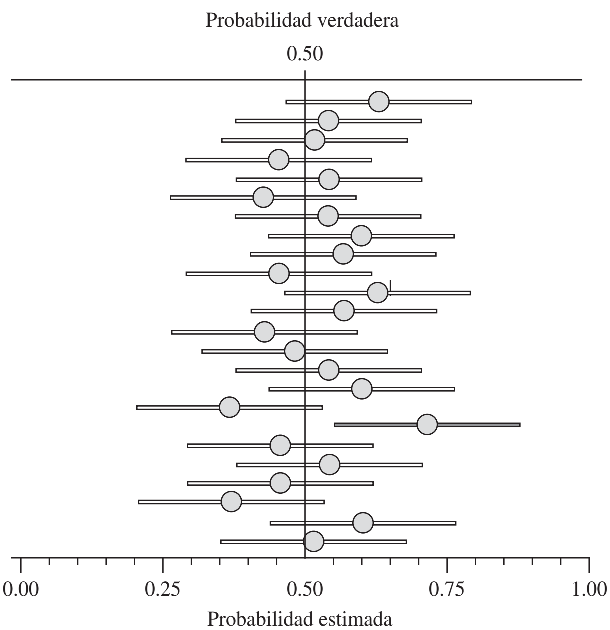
Observe que este intervalo de confianza incluye al cero. Entonces, un valor cero para la diferencia en proporciones $(p_1 - p_2)$ es "creíble" (con un nivel de confianza aproximado de 98%) con base en los datos observados. No obstante, el intervalo también incluye el valor .1. Por tanto, .1 representa otro valor de $(p_1 - p_2)$ que es "creíble" con base en los datos que hemos analizado. ■

Cerramos esta sección con una investigación práctica de la aplicación del procedimiento de estimación de intervalos de una muestra grande para una sola proporción poblacional p, con base en Y, el número de éxitos observado durante n intentos en un experimento binomial. En este caso, $\theta = p$; $\hat{\theta} = \hat{p} = Y/n$ y $\sigma_{\hat{\theta}} = \sigma_{\hat{p}} = \sqrt{p(1-p)/n} \approx \sqrt{\hat{p}(1-\hat{p})/n}$. (Al igual que en la Sección 8.3, $\sqrt{\hat{p}(1-\hat{p})/n}$ da una buena aproximación para $\sigma_{\hat{p}}$.) Los límites de confianza apropiados son entonces

$$\hat{\theta}_L = \hat{p} - z_{\alpha/2}\left[\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \quad \text{y} \quad \hat{\theta}_U = \hat{p} + z_{\alpha/2}\left[\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

La figura 8.8 muestra los resultados de 24 experimentos binomiales independientes, cada uno basado en 35 intentos cuando el verdadero valor de p es $= 0.5$. Para cada uno de los experimentos calculamos el número de éxitos y, el valor de $\hat{p} = y/35$ y el correspondiente intervalo de confianza de 95% usando la fórmula $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/35}$. (Observe que $z_{.025} = 1.96$.) En el primer experimento binomial vimos que $y = 18$, $\hat{p} = 18/35 = 0.5143$ y $\sigma_{\hat{p}} \approx \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{(.5143)(.4857)/35} = 0.0845$.. Por tanto, el intervalo obtenido en el primer experimento es $.5143 \pm 1.96(0.0845)$ o (0.3487, 0.6799). La estimación para p desde el primer experimento es mostrada por el punto grande más bajo de la Figura 8.8 y el intervalo de confianza resultante está dado por la recta horizontal que pasa por ese punto. La recta vertical indica el verdadero valor de p, 0.5 en este caso. Observe que el intervalo

FIGURA **8.8** Veinticuatro intervalos de confianza de 95% realizados para una proporción poblacional



obtenido en el primer intento (de tamaño 35) en realidad contiene el valor verdadero de la proporción poblacional p.

Los 23 intervalos de confianza restantes contenidos en esta pequeña simulación están dados por el resto de las líneas horizontales de la Figura 8.8. Observe que cada intervalo individual contiene o no el valor verdadero de p. No obstante, el valor verdadero de *p está* contenido en 23 de los 24 intervalos (95.8%) observados.

Si se utilizara muchas veces el mismo procedimiento, cada intervalo individual contendría o no el valor verdadero de p, pero el *porcentaje* de todos los intervalos que capturan p sería muy cercano a 95%. Uno tiene "95% de confianza" de que el intervalo contenga el parámetro ya que éste se obtuvo usando un *procedimiento* que genera intervalos que contienen el parámetro aproximadamente 95% de las veces que se usa.

La aplicación *ConfidenceIntervalP* (disponible en www.thomsonedu.com/statistics/wackerly) fue utilizada para generar la Figura 8.8. ¿Qué ocurre si se usan diferentes valores de n o diferentes coeficientes de confianza? ¿Obtenemos resultados similares si el valor verdadero de p es otro diferente de 0.5? Varios de los siguientes ejercicios nos permitirán usar la aplicación para contestar preguntas como éstas.

En esta sección hemos empleado el método del pivote para deducir intervalos de confianza de una muestra grande para los parámetros μ, p, $\mu_1 - \mu_2$ y $p_1 - p_2$ de acuerdo con las condiciones de la Sección 8.3. La fórmula básica es

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}},$$

donde los valores de $\hat{\theta}$ y $\sigma_{\hat{\theta}}$ aparecen en la Tabla 8.1. Cuando $\theta = \mu$ es el parámetro objetivo, entonces $\hat{\theta} = \overline{Y}$ y $\sigma^2_{\hat{\theta}} = \sigma^2/n$, donde σ^2 es la varianza poblacional. Si se conoce el valor verdadero de σ^2, debe usarse en el cálculo del intervalo de confianza; si σ^2 no se conoce y n es grande, no hay demasiada pérdida de precisión si s^2 se sustituye por σ^2 en la fórmula del intervalo de confianza. Del mismo modo, si σ_1^2 y σ_2^2 son desconocidas y tanto n_1 como n_2 son grandes, s_1^2 y s_2^2 se pueden sustituir por estos valores en la fórmula del intervalo de confianza de una muestra grande para $\theta = \mu_1 - \mu_2$.

Cuando $\theta = p$ es el parámetro objetivo, entonces $\hat{\theta} = \hat{p}$ y $\sigma_{\hat{p}} = \sqrt{pq/n}$. Como p es el parámetro objetivo desconocido, $\sigma_{\hat{p}}$ no puede ser evaluada. Si n es grande y sustituimos $\hat{p}$ por p (y $\hat{q} = 1 - \hat{p}$ por q) en la fórmula para $\sigma_{\hat{p}}$; sin embargo, el intervalo de confianza resultante tendrá aproximadamente el coeficiente de confianza establecido. Para n_1 y n_2 grandes, se cumplen enunciados semejantes cuando $\hat{p}_1$ y $\hat{p}_2$ se usan para calcular p_1 y p_2, respectivamente, en la fórmula para $\sigma^2_{\hat{p}_1 - \hat{p}_2}$. La justificación teórica para estas sustituciones se dará en la Sección 9.3.

Ejercicios

8.50 Consulte el Ejemplo 8.8. En este ejemplo, p_1 y p_2 se usaron para denotar las proporciones de refrigeradores de las marcas A y B, respectivamente, que fallaron durante los períodos de garantía.

a En el nivel aproximado de 98% de confianza, ¿cuál es el mayor "valor creíble" para la diferencia en las proporciones de fallas de refrigeradores de las marcas A y B?

b En el nivel aproximado de 98% de confianza, ¿cuál es el menor "valor creíble" para la diferencia en las proporciones de fallas de refrigeradores de las marcas A y B?

c Si p_1 - p_2 es realmente igual a 0.2251, ¿cuál marca tiene la mayor proporción de fallas durante el período de garantía? ¿Qué tanto más grande?

d Si p_1 - p_2 es realmente igual a –0.1451, ¿cuál marca tiene la mayor proporción de fallas durante el período de garantía? ¿Qué tanto más grande?

e Como se observó en el Ejemplo 8.8, cero es un valor creíble de la diferencia. ¿Concluiría usted que hay evidencia de una diferencia en las proporciones de fallas (dentro del período de garantía) para las dos marcas de refrigeradores? ¿Por qué?

8.51 **Ejercicio Applet** ¿Qué ocurre si tratamos de usar la aplicación *ConfidenceIntervalP* (disponible en www.thomsonedu.com/statistics/wackerly) para reproducir los resultados presentados en la Figura 8.8? Entre en la aplicación. No cambie el valor de p de .50 o el coeficiente de confianza de .95, pero use el botón "Sample Size" para cambiar el tamaño muestral a $n = 35$. Haga clic en el botón "One Sample" una sola vez. En la parte superior izquierda de la pantalla, los valores de la muestra están descritos por un conjunto de 35 ceros (0) y unos (1), y el valor de la estimación para p y el intervalo de confianza resultante de 95% se dan debajo de los valores de muestra.

a ¿Cuál es el valor de $\hat{p}$ que obtuvo? ¿Es igual que el primer valor obtenido, 0.5143, cuando la Figura 8.8 se generó? ¿Le sorprende esto? ¿Por qué?

b Use el valor de la estimación que obtuvo y la fórmula para un intervalo de confianza de 95% para verificar que el intervalo de confianza dado en la pantalla esté correctamente calculado.

c ¿El intervalo que usted obtuvo contiene el valor verdadero de p?

d ¿Cuál es la longitud del intervalo de confianza que usted obtuvo? ¿Es exactamente igual que la longitud del primer intervalo, (.3487, .6799), obtenida cuando se generó la Figura 8.8? ¿Por qué?

e Haga clic en el botón "One Sample" otra vez. ¿Este intervalo es diferente que el previamente generado? Haga clic en el botón "One Sample" tres veces más. ¿Cuántos intervalos claramente diferentes aparecen entre los primeros 5 intervalos generados? ¿Cuántos de ellos contienen .5?

f Haga clic en el botón "One Sample" hasta que haya obtenido 24 intervalos. ¿Qué porcentaje de los intervalos contiene el valor verdadero de $p = .5$? ¿El porcentaje es cercano al valor que usted esperaba?

8.52 **Ejercicio Applet** Consulte el Ejercicio 8.51. No cambie el valor de p de .50 o el coeficiente de confianza de .95, pero use el botón "Sample Size" para cambiar el tamaño muestral a $n = 50$. Haga clic en el botón "One Sample" una sola vez.

a ¿Qué tan largo es el intervalo de confianza resultante? ¿Cómo se compara la longitud del intervalo con la que obtuvo en el Ejercicio 8.51(d)? ¿Por qué son diferentes las longitudes de los intervalos?

b Haga clic en el botón "25 Samples". ¿El porcentaje de intervalos que contiene el valor verdadero de p es cercano a lo que usted esperaba?

c Haga clic en el botón "100 Samples". ¿El porcentaje de intervalos que contiene el valor verdadero de p es cercano a lo que usted esperaba?

d Si usted hiciera clic varias veces en el botón "100 Samples" y calculara el porcentaje de todos los intervalos que contienen el valor verdadero de p, ¿qué porcentaje de intervalos espera que contengan p?

8.53 **Ejercicio Applet** Consulte los Ejercicios 8.51 y 8.52. Cambie el valor de p a .25 (ponga el cursor en la línea vertical y arrástrelo a la izquierda hasta que aparezca .25 como la probabilidad verdadera). Cambie el tamaño muestral a $n = 75$ y el coeficiente de confianza a 0.90.

a Haga clic en el botón "One Sample" una vez.

i ¿Cuál es la longitud del intervalo resultante? ¿El intervalo es más largo o más corto que el obtenido en el Ejercicio 8.51(d)?

ii Mencione tres razones por las que el intervalo que obtuvo usted en el inciso (i) es más corto que el intervalo obtenido en el Ejercicio 8.51(d).

b Haga clic en el botón "100 Samples" unas cuantas veces. Cada clic producirá 100 intervalos y dará al usuario el número y proporción de esos 100 intervalos que contienen el valor verdadero de p. Después de cada clic anote el número de intervalos que contienen $p = .25$.

i ¿Cuántos intervalos generó? ¿Cuántos de estos intervalos contienen el verdadero valor de p?

ii ¿Qué porcentaje de todos los intervalos contienen a p?

8.54 **Ejercicio Applet** Consulte los Ejercicios 8-51–8.53. Cambie el valor de p a .90. Cambie el tamaño muestral a $n = 10$ y el coeficiente de confianza a 0.95. Haga clic en el botón "100 Samples" unas cuantas veces. Después de cada clic anote el número de intervalos que contienen $p = .90$.

a Cuando la simulación produjo diez éxitos en diez intentos, ¿cuál es el intervalo de confianza resultante de 95% para p? ¿Cuál es la longitud del intervalo? ¿Por qué? ¿Cómo se representa en la pantalla?

b ¿Cuántos intervalos generó? ¿Cuántos de los intervalos generados contienen el valor verdadero de p?

c ¿Qué porcentaje de todos los intervalos generados contienen a p?

d ¿Le sorprende el resultado del inciso c?

e ¿El resultado del inciso c invalida los procedimientos del intervalo de confianza de *muestra grande* presentados en esta sección? ¿Por qué?

8.55 **Ejercicio Applet** Consulte los Ejercicios 8.51–8.54. Cambie el valor de p a .90. Cambie el tamaño muestral a $n = 100$ y el coeficiente de confianza a .95. Haga clic en el botón "100 Samples" unas cuantas veces. Después de cada clic anote el número de intervalos que contienen $p = .90$ y conteste las preguntas planteadas en el Ejercicio 8.54, incisos b–e.

8.56 ¿Está menguando el romance de los estadounidenses con el cine? En una encuesta Gallup[5] de $n = 800$ adultos seleccionados aleatoriamente, 45% indicaron que el cine estaba mejorando mientras que 43% dijeron que el cine estaba empeorando.

a Encuentre un intervalo de confianza de 98% para p, la proporción total de adultos que dicen que el cine está mejorando.

b ¿El intervalo incluye el valor $p = 0.50$? ¿Piensa usted que una mayoría de adultos dice que el cine está mejorando?

8.57 Consulte el Ejercicio 8.29. De acuerdo con el resultado dado ahí, 51% de los $n = 1001$ adultos encuestados en noviembre de 2003 dijeron ser aficionados al beisbol. Construya un intervalo de confianza de 99% para la proporción de adultos que en noviembre de 2003 dijeron ser aficionados al beisbol (después de la Serie Mundial). Interprete este intervalo.

8.58 Los administradores de un hospital deseaban estimar el número promedio de días necesarios para el tratamiento de enfermos internados entre las edades de 25 y 34 años. Una muestra aleatoria de 500 pacientes entre estas edades produjo una media y una desviación estándar igual a 5.4 y 3.1 días, respectivamente.

5 *Fuente:* "Movie Mania Ebbing," Gallup Poll of 800 adults, http://www.usatoday.com/snapshot/news/2001-06-14-moviemania.htm., 16–18 de marzo de 2001

Construya un intervalo de confianza del 95% para la duración media de permanencia de la población de pacientes de la cual se extrajo la muestra.

8.59 Cuando se trata de anunciar, los "preadolescentes" no están listos para mensajes de línea dura que publicistas usan con frecuencia para llegar a los adolescentes. El estudio[6] del grupo Geppeto encontró que 78% de los "preadolescentes" entienden y disfrutan anuncios que son tontos por naturaleza. Suponga que el estudio comprendió $n = 1030$ "preadolescentes".

a Construya un intervalo de confianza de 90% para la proporción de "preadolescentes" que entienden y disfrutan anuncios que son tontos por naturaleza.

b ¿Piensa usted que "más de 75%" de todos los "preadolescentes" disfrutan anuncios que son tontos por naturaleza? ¿Por qué?

8.60 ¿Cuál *es* la temperatura corporal normal para personas sanas? Una muestra aleatoria de 130 temperaturas corporales en personas sanas proporcionadas por Allen Shoemaker[7] dio 98.25 grados y desviación estándar de 0.73 grados.

a Dé un intervalo de confianza de 99% para el promedio de temperatura corporal de personas sanas.

b El intervalo de confianza obtenido en el inciso a ¿contiene el valor de 98.6 grados, que es el promedio aceptado de temperatura citado por médicos y otros? ¿Qué puede usted concluir?

8.61 Una pequeña cantidad selenio, de 50 a 200 microgramos (μg) por día, es considerada esencial para una buena salud. Suponga que se seleccionaron muestras aleatorias independientes, de $n_1 = n_2 = 30$ adultos provenientes de dos regiones de Estados Unidos y que se registró para cada persona una ingesta diaria de selenio tanto de líquidos como de sólidos. La media y la desviación estándar de la ingesta diaria de selenio para los 30 adultos de la región 1 fueron $\overline{y} = 167.1$ μg y $s_1 = 24.3$ μg, respectivamente. Los estadísticos correspondientes para los 30 adultos de la región 2 fueron $\overline{y} = 140.9$ μg y $s_2 = 17.6$ μg. Encuentre un intervalo de confianza de 95% para la diferencia en la ingesta media de selenio para las dos regiones.

8.62 Los siguientes estadísticos son el resultado de un experimento realizado por P. I. Ward para investigar una teoría relativa al comportamiento de cambio de piel del macho *Gammarus pulex*, un pequeño crustáceo.[8] Si el macho cambia de piel mientras se aparea con una hembra, éste debe liberarla y perderla. La teoría es que el macho *Gammarux pulex* es capaz de posponer dicho cambio, con lo cual reduce la posibilidad de perder su pareja. Ward asignó aleatoriamente 100 parejas de machos y hembras a dos grupos de 50 cada uno. Las parejas del primer grupo se mantuvieron juntas (normal); las del segundo grupo fueron separadas. Se registró el tiempo de muda para machos y hembras, y las medias, desviaciones estándar y tamaños muestrales se ilustran en la tabla siguiente. (El número de crustáceos en cada una de las cuatro muestras es menor que 50 porque algunos en cada grupo no sobrevivieron hasta el tiempo de muda.)

	Tiempo de muda (días)		
	Media	s	n
Machos			
Normal	24.8	7.1	34
Separados	21.3	8.1	41
Hembras			
Normal	8.6	4.8	45
Separados	11.6	5.6	48

6. *Fuente:* "Caught in the Middle", *American Demographics*, julio de 2001, pp. 14-15.

7. *Fuente:* Allen L. Shoemaker, "What's Normal? Temperature, Gender and Heart Rate," *Journal of Statistics Education* (1966).

8. *Fuente:* "*Gammarus pulex* Control Their Moult Timing to Secure Mates," *Animal Behaviour* 32 (1984).

a Encuentre un intervalo de confianza de 99% para la diferencia en tiempo promedio de muda para machos "normales" contra los "separados" de sus parejas.

b Interprete el intervalo.

8.63 A la mayoría de estadounidenses les gusta participar eventos deportivos o al menos verlos. Algunos sienten que los deportes tienen más que sólo valor de entretenimiento. En una encuesta de 1000 adultos, realizada por KRC Research & Consulting, 78% sintieron que los deportes de gran atractivo tienen un efecto positivo en la sociedad.[9]

a Encuentre un intervalo de confianza de 95% para el porcentaje del público que piensa que los deportes tienen un efecto positivo en la sociedad.

b La encuesta publicó un margen de error de "más o menos 3.1%". ¿Esto concuerda con la respuesta de usted al inciso a? ¿Qué valor de p produce el margen de error dado por la encuesta?

8.64 En una encuesta de la CNN/*USA Today*/Gallup a 1000 estadounidenses se les preguntó qué tan bien los describía el término *patriótico*.[10] Algunos resultados de la encuesta están contenidos en la siguiente tabla de resumen.

	Grupo de edad		
	Todos	18–34	60+
Muy bien	.53	.35	.77
Regular	.31	.41	.17
No muy bien	.10	.16	.04
Nada bien	.06	.08	.02

a Si los grupos de 18–34 y 60 o más años estaban formados de 340 y 150 individuos, respectivamente, encuentre un intervalo de confianza de 98% para la diferencia en proporciones de los individuos en estos grupos de edades que acordaron que *patriótico* los describía muy bien.

b Con base en el intervalo que obtuvo en el inciso a, ¿piensa que la diferencia en proporciones de los que se vieron a sí mismos como patrióticos es de hasta 0.6? Explique.

8.65 Para una comparación de los porcentajes de piezas defectuosas producidas por dos líneas de montaje, de cada línea se seleccionaron muestras aleatorias independientes de 100 piezas. La línea A produjo 18 piezas defectuosas en la muestra y la línea B contenía 12 piezas defectuosas.

a Encuentre un intervalo de confianza de 98% para la verdadera diferencia en proporciones de piezas defectuosas para las dos líneas.

b ¿Hay evidencia aquí que sugiera que una línea produce una proporción más alta de piezas defectuosas que la otra?

8.66 Históricamente, la biología se ha impartido en conferencias y la evaluación de su aprendizaje se logró mediante pruebas de vocabulario y datos memorizados. Un nuevo currículo inventado por un profesor: Biología, contenido de una comunidad (BACC, por sus siglas en inglés), está basado en estándares, orientado a actividades y centrado en preguntas. Los estudiantes a quienes se enseñó a usar los métodos históricos y los nuevos fueron examinados, en el sentido tradicional, sobre conceptos de biología que destacaban conocimientos de biología y habilidades en procesos. Los resultados de un examen sobre conceptos de biología se publicaron en *The American Biology Teacher* y se muestran en la tabla siguiente.[11]

9. *Fuente:* Mike Tharp, "Ready, Set, Go. Why We Love Our Games—Sports Crazy," *U.S. News & World Report,* 15 de julio de 1997, p. 31.

10. *Fuente:* Adaptado de "I'am a Yankee Doodle Dandy," Kowledge Networks: 2000, *American Demographics,* julio de 2001, p. 9.

11. *Fuente:* William Leonard, Barbara Speziale, y John Pernick, "Performance Assessment of a Standards-Based High School Biology Curriculum," *The American Biology Teacher* 63(5) (2001):310-316.

	Media	Tamaño muestral	Desviación estándar
Pre-examen: todos los grupos BACC	13.38	372	5.59
Pre-examen: todo tradicional	14.06	368	5.45
Post-examen: todos los grupos BACC	18.50	365	8.03
Post-examen: todo tradicional	16.50	298	6.96

a Proponga un intervalo de confianza de 90% para la calificación media de postexamen para todos los estudiantes BACC.

b Encuentre un intervalo de confianza de 95% para la diferencia en las calificaciones medias de post-examen para estudiantes BACC y los que aprendieron en forma tradicional.

c ¿El intervalo de confianza del inciso b proporciona evidencia de que hay una diferencia en las calificaciones medias postexamen para estudiantes BACC y los tradicionales? Explique.

8.67 Un método sugerido para resolver la escasez de energía eléctrica en una región comprende la construcción de plantas nucleares flotantes generadoras de energía eléctrica a pocas millas de la costa en el océano. La preocupación por la posibilidad de una colisión de barcos con la planta flotante, pero anclada, ha aumentado la necesidad de una estimación de la densidad del tránsito de barcos en la zona. El número de barcos que pasan diariamente a no más de 10 millas de la ubicación propuesta de la planta eléctrica, registrada para $n = 60$ días durante julio y agosto, poseía una media muestral y varianza de $\overline{y} = 7.2$ y $s^2 = 8.8$.

a Encuentre un intervalo de confianza de 95% para el número medio de barcos que pasen a no más de 10 millas del lugar propuesto para la planta eléctrica durante un período de 1 día.

b Se esperaba que la densidad de tránsito disminuyera durante los meses de invierno. Una muestra de $n = 90$ registros diarios de avistamientos de barcos para diciembre, enero y febrero dieron una media y varianza de $\overline{y} = 4.7$ y $s^2 = 4.9$. Encuentre un intervalo de confianza de 90% para la diferencia en la densidad promedio de tránsito de barcos entre los meses de verano e invierno.

c ¿Cuál es la población asociada con la estimación en el inciso b? ¿Qué podría estar mal con el procedimiento de muestreo para los incisos a y b?

***8.68** Suponga que Y_1, Y_2, Y_3 y Y_4 tienen una distribución multinomial con n intentos y probabilidades p_1, p_2, p_3 y p_4 para las cuatro celdas. Al igual que en el caso binomial, cualquier combinación lineal de Y_1, Y_2, Y_3 y Y_4 estará distribuida normalmente en forma aproximada para n grande.

a Determine la varianza de $Y_1 - Y_2$. [*Sugerencia:* recuerde que las variables aleatorias Y_i son dependientes.]

b Un estudio de actitudes entre residentes de Florida con respecto a políticas para manejar la presencia de caimanes en zonas urbanas mostró lo siguiente. Entre 500 personas muestreadas y a las que se les dieron cuatro opciones de manejo, 6% dijeron que los caimanes deberían ser protegidos por completo, 16% dijeron que deberían ser destruidos por guardias de fauna silvestre, 52% dijeron que deberían ser reubicados vivos y 26% dijeron que debería permitirse una cosecha comercial regulada. Calcule la diferencia entre la proporción de la población que está a favor de la completa protección y la que está en pro de la destrucción por guardias de fauna silvestre. Use un coeficiente de confianza de .95.

8.69 El *Journal of Communication*, invierno de 1978, publicó un estudio acerca de ver violencia en TV. Muestras de poblaciones con bajos porcentajes de ver televisión (10-19 programas por semana) y altos porcentajes de verla (40-49 programas por semana) se dividieron en dos grupos de edades y se registró el número Y de personas que ven un alto número de programas de violencia. Los datos para dos grupos de edades se muestran en la tabla siguiente, con n_i denotando el tamaño muestral para cada celda. Si Y_1, Y_2, Y_3 y Y_4 tienen distribuciones binomiales independientes con parámetros p_1, p_2, p_3 y p_4, respectivamente, encuentre un intervalo de confianza de 95% para $(p_3 - p_1) - (p_4 - p_2)$. Esta función de los valores p_i representa una comparación entre el cambio en los hábitos de ver TV para adultos jóvenes y el cambio correspondiente para adultos mayores, cuando pasamos de los que tienen bajo porcentaje de ver TV a aquellos que tienen altos porcentajes de verla. (Los datos sugieren que el porcentaje de ver violencia puede aumentar con adultos jóvenes pero disminuye con adultos mayores.)

Porcentaje de ver TV	Grupo de edad			
	16–34		55 y más	
Bajo	$y_1 = 20$	$n_1 = 31$	$y_2 = 13$	$n_2 = 30$
Alto	$y_3 = 18$	$n_3 = 26$	$y_4 = 7$	$n_4 = 28$

8.7 Selección del tamaño muestral

El diseño de un experimento es en esencia un plan para adquirir una cantidad de información. Al igual que cualquier otra mercancía, la información se puede comprar a precios que varían dependiendo de la forma en la que se obtienen los datos. Algunas mediciones contienen una gran cantidad de información acerca del parámetro de interés; otras pueden contener poca o ninguna. La investigación, ya sea científica o de otro tipo, se hace para obtener información. Obviamente, deberíamos buscar obtener información a un costo mínimo.

El procedimiento de muestreo, o *diseño experimental*, como suele llamarse, afecta la cantidad de información por medición. Éste, junto con el tamaño muestral n controla la cantidad total de información relevante en una muestra. En esta etapa de nuestro estudio nos ocuparemos de la situación de muestreo más sencilla: muestreo aleatorio de una población relativamente grande. Primero dedicamos nuestra atención a la selección del tamaño muestral n.

Un investigador avanza poco en la planeación de un experimento antes de abordar el problema de seleccionar el tamaño muestral. De hecho, una de las preguntas más frecuentes que se plantea un estadístico es ¿cuántas mediciones deben incluirse en la muestra? Desafortunadamente, el estadístico no puede contestar esta pregunta sin saber cuánta información desea obtener el experimentador. Si nos referimos específicamente a una estimación, nos gustaría saber qué tan precisa desea el experimentador que sea. El experimentador puede indicar la precisión deseada al especificar un límite en el error de estimación.

Por ejemplo, suponga que deseamos estimar el promedio diario de producción μ de un producto químico y deseamos que el error de estimación sea menor que 5 toneladas con probabilidad de .95. Debido a que aproximadamente 95% de las medias muestrales estarán a no más de $2\sigma_{\overline{Y}}$ de μ en muestreo repetido, estamos pidiendo que $2\sigma_{\overline{Y}}$ sea igual a 5 toneladas (vea la Figura 8.9). Entonces

$$\frac{2\sigma}{\sqrt{n}} = 5 \quad \text{y} \quad n = \frac{4\sigma^2}{25}.$$

No podemos obtener un valor numérico exacto para n a menos que se conozca la desviación estándar poblacional σ. Esto es exactamente lo que esperaríamos porque la variabilidad asociada con el estimador $\overline{Y}$ depende de la variabilidad exhibida en la población de la cual se sacó la muestra.

A falta de un valor exacto para σ, usamos la mejor aproximación disponible, por ejemplo una estimación s obtenida de una muestra previa, o el conocimiento de la amplitud de las mediciones en la población. Como la amplitud es aproximadamente igual a 4σ (recuerde la regla empírica), un cuarto de la amplitud da un valor aproximado de σ. Para nuestro ejemplo

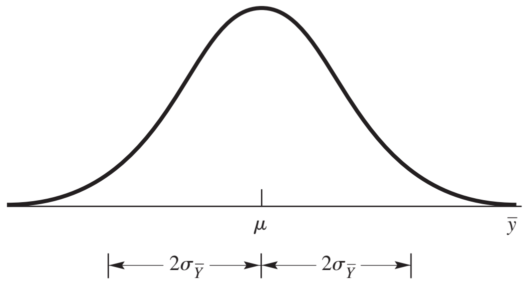


FIGURA 8.9 Distribución aproximada de $\overline{Y}$ para muestras grandes

suponga que se sabe que la amplitud de la producción es aproximadamente de 84 toneladas. Entonces $\sigma \approx 84/4 = 21$ y

$$n = \frac{4\sigma^2}{25} \approx \frac{(4)(21)^2}{25} = 70.56$$
$$= 71.$$

Usando un tamaño muestral $n = 71$, podemos tener cierta seguridad (con un coeficiente de confianza de alrededor de .95) de que nuestro cálculo se encuentra a no más de 5 toneladas del verdadero promedio diario de producción.

En realidad, esperaríamos que el error de estimación fuera mucho menor que 5 toneladas. De acuerdo con la regla empírica, la probabilidad es aproximadamente igual a .68 de que el error de estimación sea menor que $\sigma_{\overline{Y}} = 2.5$ toneladas. Las probabilidades .95 y .68 empleadas en estos enunciados son inexactas porque σ fue aproximada. Aun cuando este método de seleccionar el tamaño muestral es sólo aproximado para una precisión especificada de estimación, es el mejor del que se dispone y es ciertamente mejor que seleccionar el tamaño muestral de manera intuitiva.

El método de seleccionar los tamaños muestrales para todos los procedimientos de estimación de muestra grande indicados en la Tabla 8.1 es análogo al que acabamos de describir. El experimentador debe especificar un límite deseado en el error de estimación y un nivel de confianza asociado $1 - \alpha$. Por ejemplo, si el parámetro es θ y el límite deseado es B, igualamos

$$z_{\alpha/2}\sigma_{\hat{\theta}} = B,$$

donde, como en la Sección 8.6,

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2}.$$

Ilustramos el uso de este método en los siguientes ejemplos.

EJEMPLO **8.9** La reacción de un individuo a un estímulo en un experimento psicológico puede tomar una de dos formas, A o B. Si un experimentador desea estimar la probabilidad p de que una persona reaccione en una forma A, ¿cuántas personas deben incluirse en el experimento? Suponga que el experimentador estará satisfecho si el error de estimación es menor que .04 con probabilidad igual a .90. Suponga también que él espera que p se encuentre en algún punto cercano a .6.

Solución Debido a que hemos especificado que $1 - \alpha = .90$, α debe ser igual a .10 y $\alpha/2 = .05$. El valor z correspondiente a un área igual a .05 en la cola superior de la distribución normal estándar es $z_{\alpha/2} = z_{.05} = 1.645$. Entonces requerimos que

$$1.645\sigma_{\hat{p}} = .04, \quad \text{o} \quad 1.645\sqrt{\frac{pq}{n}} = .04.$$

Como el error estándar de $\hat{p}$ depende de p, que es desconocido, podríamos usar el valor estimado de $p = .6$ dado por el experimentador como un valor aproximado para n. Entonces

$$1.645\sqrt{\frac{(.6)(.4)}{n}} = .04$$
$$n = 406.$$

En este ejemplo supusimos que $p \approx .60$. ¿Cómo procederíamos si no tuviéramos idea del valor verdadero de p? En el Ejercicio 7.76(a) establecimos que el valor *máximo* para la varianza de $\hat{p} = Y/n$ se presenta cuando $p = .5$. Si no supiéramos que $p \approx .6$, usaríamos $p = .5$, que daría el máximo valor posible para n: $n = 423$. No importa cuál sea el verdadero valor de p, $n = 423$ es lo suficientemente grande para dar una estimación que esté a no más de $B = .04$ de p con probabilidad .90. ■

EJEMPLO **8.10** Un experimentador desea comparar la efectividad de dos métodos de capacitación para obreros que van a realizar una operación de ensamble. Los obreros seleccionados han de dividirse en dos grupos de igual tamaño, el primero para recibir el método 1 de capacitación y el segundo el método 2 de capacitación. Después de la capacitación cada obrero realizará la operación de ensamble y se registrará el tiempo que le tome hacerlo. El experimentador espera que las mediciones para ambos grupos tengan una amplitud de aproximadamente 8 minutos. Si la estimación de la diferencia en los tiempos promedio de ensamble debe ser correcta con una variación de no más de 1 minuto con probabilidad .95, ¿cuántos trabajadores deben incluirse en cada grupo de capacitación?

Solución El fabricante especificó $1 - \alpha = .95$. Entonces, $\alpha = .05$ y $z_{\alpha/2} = z_{.025} = 1.96$. Igualando $1.96\sigma_{(\overline{Y}_1 - \overline{Y}_2)}$ a 1 minuto, obtenemos

$$1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1.$$

De manera alternativa, como deseamos que n_1 sea igual a n_2, podemos hacer $n_1 = n_2 = n$ y obtener la ecuación

$$1.96\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} = 1.$$

Como ya dijimos antes, la variabilidad de cada método de ensamble es aproximadamente igual; en consecuencia, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Debido a que la amplitud, 8 minutos, es aproximadamente igual a 4σ, tenemos

$$4\sigma \approx 8, \text{ o bien, lo que es equivalente, } \sigma \approx 2.$$

Sustituyendo este valor por σ_1 y σ_2 en la ecuación anterior, obtenemos

$$1.96\sqrt{\frac{(2)^2}{n}+\frac{(2)^2}{n}}=1.$$

Resolviendo, obtenemos $n = 30.73$. Por tanto, cada grupo debe contener $n = 31$ miembros. ■

Ejercicios

8.70 Sea Y una variable aleatoria binomial con parámetro p. Encuentre el tamaño muestral necesario para calcular p con tolerancia de no más de .05 con probabilidad .95 en las siguientes situaciones:

a Si se considera que p es aproximadamente .9.

b Si no se conoce información acerca de p (use $p = .5$ para calcular la varianza de $\hat{p}$).

8.71 Un servicio estatal de fauna silvestre desea calcular el número promedio de días que cada cazador con licencia se dedica a esta actividad realmente durante una estación determinada, con un límite en el error de estimación igual a 2 días de caza. Si los datos recolectados en estudios anteriores han demostrado que σ es aproximadamente igual a 10, ¿cuántos cazadores deben estar incluidos en el estudio?

8.72 Es frecuente que encuestadores por teléfono entrevisten entre 1000 y 1500 personas sobre sus opiniones en asuntos varios. ¿El rendimiento de los equipos de atletismo universitarios tiene un impacto positivo en la percepción del público del prestigio de las instituciones? Una nueva encuesta se va a efectuar para ver si hay diferencia entre las opiniones de hombres y mujeres sobre este asunto.

a Si se han de entrevistar 1000 hombres y 1000 mujeres, ¿con cuánta precisión podría usted estimar la diferencia en las proporciones que piensan que el rendimiento de sus equipos de atletismo tiene un impacto positivo en la percepción del público acerca del prestigio de las instituciones? Encuentre un límite para el error de estimación.

b Supongamos que usted estuviera diseñando la encuesta y desea estimar la diferencia en un par de proporciones, correcta a no más de .02, con probabilidad .9. ¿Cuántas entrevistas deben incluirse en cada muestra?

8.73 Consulte el Ejercicio 8.59. ¿Cuántos "preadolescentes" deben haber sido entrevistados para calcular la proporción de ellos que entienden y disfrutan de anuncios de naturaleza tonta, correcto a no más de .02, con probabilidad 0.99? Use la proporción del ejemplo anterior para calcular el error estándar de la estimación.

8.74 Supongamos que usted desea calcular el pH promedio de precipitaciones en una zona que sufre de fuerte contaminación debido a la descarga de humo de una planta de energía eléctrica. Suponga que σ está en las cercanías de .5 pH y que se desea que el cálculo varíe en no más de .1 de μ con probabilidad cercana a .95. ¿Aproximadamente cuántas precipitaciones deben estar incluidas en la muestra (una lectura de pH por precipitación)? ¿Sería válido seleccionar todos las muestras de agua de una sola precipitación? Explique.

8.75 Consulte el Ejercicio 8.74. Supongamos que usted desea calcular la diferencia entre la acidez media para lluvias en dos lugares diferentes, uno en una zona relativamente poco contaminada a lo largo del océano y la otra en una región sometida a fuerte contaminación del aire. Si se desea que el cálculo sea correcto al .1 pH más cercano con probabilidad cercana a .90, ¿aproximadamente cuántas precipitaciones (valores de pH) deben incluirse en cada muestra? (Suponga que la varianza de las mediciones de pH es alrededor de .25 en ambos lugares y que las muestras han de ser del mismo tamaño.)

8.76 Consulte la comparación de la ingesta diaria de selenio para un adulto en dos diferentes regiones de Estados Unidos del Ejercicio 8.61. Suponga que se desea calcular la diferencia en la ingesta diaria promedio entre las dos regiones, con un error máximo de no más de 5 μg, con probabilidad .90. Si se planea seleccionar un número igual de adultos de las dos regiones (esto es, si $\mu_1 = \mu_2$), ¿qué tan grandes deben ser n_1 y n_2?

8.77 Consulte el Ejercicio 8.28. Si el investigador desea calcular la diferencia en proporciones a no más de .05 con 90% de confianza, ¿cuántos egresados y no egresados deben ser entrevistados? (Suponga que será entrevistado un número igual de cada grupo.)

8.78 Consulte el Ejercicio 8.65. ¿Cuántas piezas deben muestrearse de cada línea si un intervalo de confianza de 95% para la diferencia real entre las proporciones ha de tener un ancho de .2? Suponga que muestras de igual tamaño se tomarán de cada línea.

8.79 Consulte el Ejercicio 8.66.

a Se ha de emprender otro estudio semejante para comparar las calificaciones promedio de postexamen para estudiantes de biología de secundaria BACC y los que aprenden de manera tradicional. El objetivo es producir un intervalo de confianza de 99% para la diferencia real en las calificaciones promedio de postexamen. Si necesitamos muestrear un número igual de estudiantes de BACC y de los que aprenden de manera tradicional y buscamos que el ancho del intervalo de confianza sea 1.0, ¿cuántas observaciones deben incluirse en cada grupo?

b Repita los cálculos del inciso a si estamos interesados en comparar calificaciones promedio de preexamen.

c Suponga que la investigadora desea construir intervalos de confianza de 99% para comparar *ambas* calificaciones, las de preexamen y las de postexamen, para estudiantes de biología de BACC y los que aprenden de manera tradicional. Si el objetivo de la investigadora es que ambos intervalos tengan anchos no mayores que 1 unidad, ¿qué tamaños muestrales deben usarse?

8.8 Intervalos de confianza de una muestra pequeña para μ y $\mu_1 - \mu_2$

Los intervalos de confianza para una media poblacional μ, que estudiaremos en esta sección, están basados en la suposición de que la muestra del experimentador se ha seleccionado aleatoriamente de entre una población normal. Los intervalos son apropiados para muestras de cualquier tamaño y los coeficientes de confianza de los intervalos son cercanos a los valores especificados aun cuando la población no sea normal, mientras la desviación no sea excesiva. Raras veces conocemos la forma de la distribución de frecuencia poblacional antes de muestrear y, en consecuencia, si un estimador de intervalo debe ser de cualquier valor, debe funcionar razonablemente bien aun cuando la población no sea normal. "Funcionar bien" significa que el coeficiente de confianza no debe ser afectado por desviaciones pequeñas de la normalidad. Para la mayor parte de las distribuciones poblacionales en forma de campana, los estudios experimentales indican que estos intervalos de confianza mantienen coeficientes de confianza cercanos a los valores nominales empleados en su cálculo.

Suponemos que $Y_1, Y_2, \ldots, Y_n$ representa una muestra aleatoria seleccionada de una población normal y con $\overline{Y}$ y S^2 representamos la media y la varianza muestrales, respectivamente. Nos gustaría construir un intervalo de confianza para la media poblacional, cuando $V(Y_i) = \sigma^2$ sea desconocida y el tamaño de la muestra sea demasiado pequeño para permitirnos aplicar las técnicas de muestra grande expuestas en la sección anterior. Dadas las suposiciones que

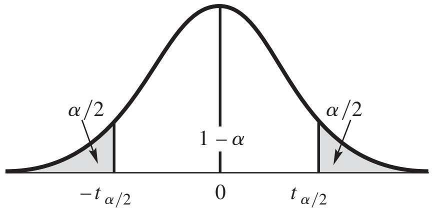


FIGURA **8.10** Ubicación de $t_{\alpha/2}$ y $-t_{\alpha/2}$

acabamos de indicar, los Teoremas 7.1 y 7.3 y la Definición 7.2 implican que

$$T = \frac{Y - \mu}{S/\sqrt{n}}$$

tiene una distribución t con $(n - 1)$ grado de libertad. La cantidad T sirve como cantidad pivote que usaremos para formar un intervalo de confianza para μ. De la Tabla 5, Apéndice 3, podemos hallar valores $t_{\alpha/2}$ y $-t_{\alpha/2}$ (vea la Figura 8.10) de modo que

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha.$$

La distribución t tiene una función de densidad muy semejante a la densidad normal estándar excepto que las colas son más anchas (como se ilustra en la Figura 7.3). Recuerde que los valores de $t_{\alpha/2}$ dependen de los grados de libertad $(n - 1)$ así como del coeficiente de confianza $(1 - \alpha)$.

El intervalo de confianza para μ se obtiene al manipular las desigualdades del enunciado de probabilidad, de modo análogo al empleado en la deducción presentada en el Ejemplo 8.6. En este caso, el intervalo de confianza resultante para μ es

$$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right).$$

De acuerdo con las suposiciones anteriores, también podemos obtener límites de confianza *unilaterales* de $100(1 - \alpha)\%$ para μ. Observe que t_α, dada en la Tabla 5, Apéndice 3, es tal que

$$P(T \leq t_\alpha) = 1 - \alpha.$$

Sustituyendo T en esta expresión y manipulando la desigualdad resultante obtenemos

$$P[\overline{Y} - t_\alpha(S/\sqrt{n}) \leq \mu] = 1 - \alpha.$$

Entonces, $\overline{Y} - t_\alpha(S/\sqrt{n})$ es un *límite de confianza inferior* de $100(1 - \alpha)\%$ para μ. De la misma manera $\overline{Y} + t_\alpha(S/\sqrt{n})$ es un límite de confianza superior de $100(1 - \alpha)\%$ para μ. Al igual que en el caso de una muestra grande, si determinamos *los* límites de confianza superior e inferior $100(1 - \alpha)\%$ para μ y usamos los límites respectivos como puntos extremos para un intervalo de confianzas el intervalo *bilateral* resultante tiene coeficiente de confianza igual a $1 - 2\alpha$.

EJEMPLO **8.11** Un fabricante ha inventado una nueva pólvora que fue probada en ocho proyectiles. Las velocidades resultantes en la boca del cañón, en pies por segundo, fueron las siguientes:

3005	2925	2935	2965
2995	3005	2937	2905

Encuentre un intervalo de confianza de 95% para el verdadero promedio de velocidad μ para proyectiles de este tipo. Suponga que las velocidades en la boca del cañón están distribuidas normalmente en forma aproximada.

Solución Si suponemos que las velocidades Y_i están distribuidas normalmente, el intervalo de confianza para μ es

$$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right),$$

donde $t_{\alpha/2}$ está determinado por un $n - 1$ grados de libertad. Para los datos dados, $\overline{y} = 2959$ y $s = 39.1$. En este ejemplo tenemos $n - 1 = 7$ grados de libertad y usando la Tabla 5, Apéndice 3, $t_{\alpha/2} = t_{.025} = 2.365$. Entonces, obtenemos

$$2959 \pm 2.365\left(\frac{39.1}{\sqrt{8}}\right), \quad \text{o bien,} \quad 2959 \pm 32.7,$$

como el intervalo de confianza observado para μ. ■

Suponga que estamos interesados en comparar las medias de dos poblaciones normales, una con media μ_1 y varianza σ_1^2 y la otra con media μ_2 y varianza σ_2^2. Si las muestras son independientes, los intervalos de confianza para $\mu_1 - \mu_2$ basados en una variable aleatoria con distribución t se pueden construir si suponemos que las dos poblaciones tienen una varianza común pero desconocida $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (desconocida).

Si $\overline{Y}_1$ y $\overline{Y}_2$ son las medias muestrales respectivas obtenidas de muestras aleatorias independientes de poblaciones normales, el intervalo de confianza de muestra grande para $(\mu_1 - \mu_2)$ se desarrolla usando

$$Z = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

como una cantidad pivote. Como supusimos que las poblaciones muestreadas están distribuidas normalmente, Z tiene una distribución normal estándar y usando la suposición $\sigma_1^2 = \sigma_2^2 = \sigma^2$, la cantidad Z se puede reescribir como

$$Z = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

Debido a que σ es desconocida, necesitamos hallar un estimador de la varianza común σ^2 para que podamos generar una cantidad con distribución t.

Denotemos con $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ la muestra aleatoria de tamaño n_1 de la primera población y denotemos con $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ una muestra aleatoria independiente de tamaño n_2 de la segunda población. Entonces

$$\overline{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i} \quad \text{y} \quad \overline{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}.$$

El estimador insesgado usual de la varianza común σ^2 se obtiene al agrupar los datos muestrales para obtener el *estimador ponderado* S_p^2:

$$S_p^2 = \frac{\sum_{i=1}^{n_1}(Y_{1i} - \overline{Y}_1)^2 + \sum_{i=1}^{n_2}(Y_{2i} - \overline{Y}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

donde S_i^2 es la varianza muestral de la i-ésima muestra, $i = 1, 2$. Observe que si $n_1 = n_2$, S_p^2 es simplemente el promedio de S_1^2 y S_2^2. Si $n_1 \neq n_2$, S_p^2 es el promedio *ponderado* de S_1^2 y S_2^2, con mayor peso asignado a la varianza muestral asociada con el tamaño muestral más grande. Además,

$$W = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} = \frac{\sum_{i=1}^{n_1}(Y_{1i} - \overline{Y}_1)^2}{\sigma^2} + \frac{\sum_{i=1}^{n_2}(Y_{2i} - \overline{Y}_2)^2}{\sigma^2}$$

es la suma de dos variables aleatorias independientes con distribución χ^2 y con $(n_1 - 1)$ y $(n_2 - 1)$ grados de libertad, respectivamente. Entonces, W tiene una distribución χ^2 con $\nu = (n_1 - 1) + (n_2 - 1) = (n_1 + n_2 - 2)$ grados de libertad. (Vea Teoremas 7.2 y 7.3.) Ahora usamos la variable W con distribución χ^2 y la cantidad Z normal, estándar, *independiente* del párrafo anterior para formar una cantidad pivote:

$$T = \frac{Z}{\sqrt{\frac{W}{\nu}}} = \left[\frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right] \Bigg/ \sqrt{\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2(n_1 + n_2 - 2)}}$$

$$= \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

una cantidad que por construcción tiene una distribución t con $(n_1 + n_2 - 2)$ grados de libertad.

Procediendo como hicimos antes en esta sección, vemos que el intervalo de confianza para $(\mu_1 - \mu_2)$ tiene la forma

$$(\overline{Y}_1 - \overline{Y}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

donde $t_{\alpha/s}$ se determina a partir de ls distribución t con $(n_1 + n_2 - 2)$ grados de libertad.

EJEMPLO **8.12** Para alcanzar la máxima eficiencia al realizar una operación de ensamble en una planta manufacturera, obreros nuevos requieren aproximadamente un periodo de capacitación de 1 mes. Se sugirió un nuevo método de capacitación y se realizó un examen para comparar el nuevo método contra el procedimiento estándar. Dos grupos de nueve obreros nuevos cada uno fueron capacitados durante 3 semanas, un grupo usando el nuevo método y el otro siguiendo el procedimiento estándar de capacitación. El tiempo (en minutos) requerido por cada obrero

Tabla 8.3 Datos para el Ejemplo 8.12

Procedimiento	Mediciones								
Estándar	32	37	35	28	41	44	35	31	34
Nuevo	35	31	29	25	34	40	27	32	31

para ensamblar el dispositivo se registró al final del período de 3 semanas. Las mediciones resultantes son las que se muestran en la Tabla 8.3. Calcule la diferencia real de las medias $(\mu_1 - \mu_2)$ con coeficiente de confianza .95. Suponga que los tiempos de ensamble están distribuidos normalmente en forma aproximada, que las varianzas de los tiempos de ensamble son aproximadamente iguales para los dos métodos y que las muestras son independientes.

Solución Para los datos de la Tabla 8.3, con la muestra 1 denotando el procedimiento estándar, tenemos

$$\begin{aligned} \overline{y}_1 &= 35.22, & \overline{y}_2 &= 31.56, \\ \sum_{i=1}^{9}(y_{1i} - \overline{y}_1)^2 &= 195.56, & \sum_{i=1}^{9}(y_{2i} - \overline{y}_2)^2 &= 160.22, \\ s_1^2 &= 24.445, & s_2^2 &= 20.027. \end{aligned}$$

Por tanto,

$$s_p^2 = \frac{8(24.445) + 8(20.027)}{9 + 9 - 2} = \frac{195.56 + 160.22}{16} = 22.236 \quad \text{y} \quad s_p = 4.716.$$

Observe que, como $n_1 = n_2 = 9$, s_p^2 es el promedio simple de s_1^2 y s_1^2. También, $t_{.025} = 2.120$ para $(n_1 + n_2 - 2) = 16$ grados de libertad. El intervalo de confianza observado es por tanto

$$(\overline{y}_1 - \overline{y}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(35.22 - 31.56) \pm (2.120)(4.716)\sqrt{\frac{1}{9} + \frac{1}{9}}$$

$$3.66 \pm 4.71.$$

Este intervalo de confianza se puede escribir en la forma [-1.05, 8.37]. El intervalo es bastante ancho e incluye valores positivos y negativos. Si $\mu_1 - \mu_2$ es positivo, $\mu_1 > \mu_2$ y el procedimiento estándar tiene un tiempo de ensamble esperado mayor que el nuevo procedimiento. Si $\mu_1 - \mu_2$ es realmente negativo, lo inverso es verdadero. Como el intervalo contiene valores positivos y negativos, se puede decir que ninguno de los métodos de capacitación produce un tiempo medio de ensamble que difiera del otro. ■

Resumen de intervalos de confianza de muestra pequeña para medias de distribuciones normales con varianza(s) desconocida(s)

Parámetro	*Intervalo de confianza* (ν = *grados de libertad*)
μ	$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right), \qquad \nu = n - 1.$
$\mu_1 - \mu_2$	$(\overline{Y}_1 - \overline{Y}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$
	donde $\nu = n_1 + n_2 - 2$ y
	$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
	(requiere que las muestras sean independientes y la suposición de que $\sigma_1^2 = \sigma_2^2$).

Cuando el tamaño (o tamaños) de la muestra se hace grande, el número de grados de libertad para la distribución *t* aumenta y la distribución *t* puede ser calculada en forma muy precisa por la distribución normal estándar. En consecuencia, los intervalos de confianza de una muestra pequeña de esta sección son casi indistinguibles respecto de los intervalos de confianza de muestra grande de la Sección 8.6 para *n* grande (o n_1 y n_2 grandes). Los intervalos son casi equivalentes cuando los grados de libertad exceden de 30.

Los intervalos de confianza para una sola media y la diferencia en dos medias se generaron con las suposiciones de que las poblaciones de interés están distribuidas normalmente. Hay suficiente evidencia práctica de que estos intervalos mantienen su coeficiente de confianza nominal mientras las poblaciones muestreadas tengan distribuciones con forma casi de campana. Si $n_1 \approx n_2$, los intervalos de $\mu_1 - \mu_2$ también mantienen sus coeficientes de confianza nominales mientras las varianzas poblacionales sean aproximadamente iguales. La independencia de las muestras es la suposición más crítica al usar los intervalos de confianza desarrollados en esta sección para comparar dos medias poblacionales.

Ejercicios

8.80 Aun cuando hay muchos tratamientos para la *bulimia nervosa*, algunas personas no se benefician de ellos. En un estudio para determinar qué factores predicen quién se beneficiará con el tratamiento, Wendy Baell y E. H. Wertheim[12] encontraron que la autoestima era uno de los pronosticadores importantes. La media y la desviación estándar de los valores postratamiento de autoestima para $n = 21$ personas fueron $\overline{y} = 26.6$ y $s = 7.4$, respectivamente. Encuentre un intervalo de confianza de 95% para los verdaderos valores de postratamiento de autoestima.

8.81 Las longitudes de caparazones de diez langostas examinados en un estudio de la infestación de la langosta *Thenus orientalis* por dos tipos de lapas, *Octolasmis tridens* y *O. lowei* se dan en la siguiente tabla.

12. *Fuente:* Wendy K. Baell y E. H. Wertheim, "Predictors of Outcome in the Treatment of Bulimia Nervosa," *British Journal of Clinical Psychology* 31 (1992).

Encuentre un intervalo de confianza de 95% para la longitud media de caparazón (en milímetros, mm) de langostas *T. orientalis* atrapadas en los mares en las cercanías de Singapur.[13]

Número de campo de langosta	*A061*	*A062*	*A066*	*A070*	*A067*	*A069*	*A064*	*A068*	*A065*	*A063*
Longitud del caparazón (mm)	78	66	65	63	60	60	58	56	52	50

8.82 Las calificaciones del Examen de Evaluación Escolar (SAT por sus siglas en inglés), que han bajado lentamente desde el inicio del examen, ahora han empezado a subir. Originalmente, una calificación de 500 estaba considerada como promedio. Las calificaciones medias para 2005 fueron aproximadamente 508 para el examen verbal y 520 para el examen de matemáticas. Una muestra aleatoria de las calificaciones del examen, de 20 alumnos de último año de una preparatoria urbana de gran tamaño, produjo las medias y desviaciones estándar citadas en la tabla siguiente:

	Verbal	Matemáticas
Media muestral	505	495
Desviación estándar muestral	57	69

a Encuentre un intervalo de confianza de 90% para la media de calificaciones del SAT verbal para alumnos de último año de preparatoria urbana.

b ¿El intervalo hallado por usted en el inciso a incluye el valor 508, la calificación media real del SAT verbal para 2005? ¿Qué puede concluir?

c Construya un intervalo de confianza de 90% para la calificación media del SAT de matemáticas para alumnos de último año de preparatoria urbana. ¿El intervalo incluye 520, la calificación media real de matemáticas para 2005? ¿Qué puede concluir?

8.83 El síndrome crónico de la sección anterior es un estado de salud caracterizado por dolor inducido por ejercicio en la parte inferior de las piernas. Hinchazón y una función deteriorada de nervios y músculos también acompañan al dolor, que se alivia con reposo. Susan Beckham y sus colegas[14] realizaron un experimento que abarcó diez corredores en buenas condiciones físicas, así como diez ciclistas también en buenas condiciones físicas, para determinar si las mediciones de presión dentro de la sección anterior del músculo difieren entre corredores y ciclistas. Los datos —presión en la sección, en milímetros de mercurio— se resumen en la tabla siguiente:

	Corredores		Ciclistas	
Condición	Media	s	Media	s
Reposo	14.5	3.92	11.1	3.98
80% de máximo consumo de O_2	12.2	3.49	11.5	4.95

a Construya un intervalo de confianza de 95% para la diferencia en la media de las presiones en la sección entre corredores y ciclistas en condición de reposo.

b Construya un intervalo de confianza de 90% para la diferencia en la media de las presiones en la sección entre corredores y ciclistas que se ejercitan al 80% de máximo consumo de oxígeno (O_2).

c Considere los intervalos construidos en los incisos a y b. ¿Cómo interpretaría los resultados obtenidos?

13. *Fuente:* W. B. Jeffries, H. K. Voris y C. M.Yang, “Diversity and Distribution of the Pedunculate Barnacle *Octolasmis* Gray, 1825 Epizoic on the Scyllarid Lobster, *Thenus orientalis* (Lund 1793),” *Crustaceana* 46(3)(1984).

14. *Fuente:* S. J. Beckham, W. A. Grana, P. Buckley, J. E. Breasile, y P. L. Claypool, “A Comparison of Anterior Compartment Pressures in Competitive Runners and Cyclists,” *American Journal of Sports Medicine* 21(1)(1993).

8.84 Es frecuente que los químicos orgánicos purifiquen compuestos orgánicos por medio de un método conocido como cristalización fraccional. Un experimentador desea preparar y purificar 4.85 g de anilina. Diez especímenes de 4.85 gramos de anilina se prepararon y purificaron para producir acetanilida. Se obtuvieron los siguientes resultados en seco:

3.85, 3.88, 3.90, 3.62, 3.72, 3.80, 3.85, 3.36, 4.01, 3.82

Construya un intervalo de confianza de 95% para el número medio de gramos de acetanilida que se puede recuperar de 4.85 gramos de anilina.

8.85 Dos nuevos medicamentos se dieron a pacientes con hipertensión. El primero de ellos bajó la presión sanguínea de 16 pacientes un promedio de 11 puntos, con una desviación estándar de 6 puntos; el segundo bajó la presión de otros 20 pacientes en un promedio de 12 puntos, con desviación estándar de 8 puntos. Determine un intervalo de confianza de 95% para la diferencia en las reducciones medias en presión sanguínea, suponiendo que las mediciones están distribuidas normalmente con varianzas iguales.

8.86 ¿El precio pagado por el atún depende del método de empaque? *Consumer Reports* da el precio promedio estimado para una lata de 6 onzas o una bolsa de 7.06 onzas de atún, con base en precios pagados a nivel nacional en supermercados.[15] Los precios se registran para una variedad de marcas de atún en la tabla siguiente:

Atún claro en agua		Atún blanco en aceite	Atún blanco en agua		Atún claro en aceite	
0.99	0.53	1.27	1.49	1.29	2.56	0.62
1.92	1.41	1.22	1.29	1.00	1.92	0.66
1.23	1.12	1.19	1.27	1.27	1.30	0.62
0.85	0.63	1.22	1.35	1.28	1.79	0.65
0.65	0.67				1.23	0.60
0.69	0.60					0.67
0.60	0.66					

Suponga que las marcas de atún incluidas en el estudio representan una muestra aleatoria de todas las marcas de atún existentes en Estados Unidos. Encuentre un intervalo de confianza de 95%

a para el precio promedio de atún claro empacado en agua. Interprete el intervalo. Específicamente, ¿a qué se refiere el “95%”?,

b para el precio promedio de atún claro empacado en aceite. ¿Cómo se compara el ancho de este intervalo con el del intervalo hallado en el inciso a? Dé tres razones por las que difieren las longitudes de los intervalos.

8.87 Consulte el Ejercicio 8.86.

a Construya un intervalo de confianza de 90% para la diferencia en el precio medio de atún claro empacado en agua y atún claro empacado en aceite.

b Con base en el intervalo obtenido en el inciso a, ¿piensa usted que los precios medios difieren para atún claro empacado en agua y aceite? ¿Por qué?

8.88 La Environmental Protection Agency (EPA) ha recolectado datos sobre mediciones de LC50 (concentraciones que matan a 50% de los animales de prueba) para ciertos productos químicos que es probable se

15. *Fuente:* Caso real “Pricing of Tuna” Copyright 2001 por la Consumers Union of U.S.,Inc., Yonkers,N.Y. 1073-1057, organización sin fines de lucro. De la edición de junio de 2001 de *Consumers Reports* © sólo para fines educacionales. NO se permite el uso comercial ni la reproducción. www.ConsumerReports.org.

encuentren en ríos y lagos de agua dulce. (Para más detalles, vea el Ejercicio 7.13.) Para cierta especie de peces, las mediciones de LC50 (en partes por millón) de DDT en 12 experimentos fueron las siguientes:

16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9

Calcule la media real de LC50 para DDT con un coeficiente de confianza 0.90. Suponga que las mediciones de LC50 tienen una distribución aproximadamente normal.

8.89 Consulte el Ejercicio 8.88. Otro insecticida común, el diazinón, dio mediciones de LC50 en tres experimentos de 7.8, 1.6 y 1.3.

a Calcule la media de LC50 para diazinón, con un intervalo de confianza de 90%.

b Calcule la diferencia entre la media de LC50 para DDT y para diazinón, con un intervalo de confianza de 90%. ¿Qué suposiciones son necesarias para que el método que se usó sea válido?

8.90 ¿Las calificaciones del SAT para estudiantes de preparatoria difieren dependiendo del campo de estudio futuro de los estudiantes? Quince estudiantes que deseaban especializarse en ingeniería se compararon con 15 estudiantes que deseaban especializarse en idioma y literatura. En la siguiente tabla se dan las medias y desviaciones estándar de las calificaciones de la parte verbal y de matemáticas de los exámenes SAT para los dos grupos de estudiantes: [16]

	Verbal		Matemáticas	
Ingeniería	$\overline{y} = 446$	$s = 42$	$\overline{y} = 548$	$s = 57$
Idiomas/literatura	$\overline{y} = 534$	$s = 45$	$\overline{y} = 517$	$s = 52$

a Construya un intervalo de confianza de 95%, para la diferencia en el promedio de calificaciones de examen verbal de estudiantes que se especializan en ingeniería y los que se especializan en idiomas/literatura.

b Construya un intervalo de confianza para la diferencia en el promedio de calificaciones de matemáticas para estudiantes que se especializan en ingeniería y para los que se especializan en idiomas/literatura.

c Interprete los resultados obtenidos en los incisos a y b.

d ¿Qué suposiciones son necesarias para que sean válidos los métodos empleados previamente?

8.91 Biólogos de la Comisión de Caza y Pesca de Florida observaron las zonas de distribución estacionales (en hectáreas) para caimanes en un lago en las afueras de Gainesville, Florida. Cinco caimanes observados en la primavera mostraron zonas de distribución de 8.0, 12.1, 8.1, 18.2 y 31.7. Cuatro caimanes diferentes observados en el verano mostraron zonas de distribución de 102.0, 81.7, 54.7 y 50.7. Calcule la diferencia entre zonas de distribución medias en primavera y verano, con un intervalo de confianza de 95%. ¿Qué suposiciones hizo?

8.92 El cobre sólido, producido por sinterización (calentamiento sin fundir) de un polvo en condiciones ambientales especificadas, se mide a continuación para ver su porosidad (en fracción de volumen debido a huecos) en un laboratorio. Una muestra de $n_1 = 4$ mediciones independientes de porosidad tienen una media de $\overline{y}_1 = .22$ y varianza de $s_1^2 = .0010.$. Un segundo laboratorio repite el mismo proceso en cobre sólido formado de un polvo idéntico y obtiene $n_2 = 5$ mediciones independientes de porosidad con $\overline{y}_2 = .17$ y $s_2^2 = .0020.$ Calcule la diferencia real entre las medias poblacionales $(\mu_1 - \mu_2)$ para estos dos laboratorios, con un coeficiente de confianza de .95.

***8.93** Una fábrica opera con dos máquinas de tipo A y una máquina de tipo B. Los costos X de reparaciones semanales para máquinas tipo A están normalmente distribuidos con media μ_1 y varianza σ^2. Los costos de reparaciones semanales Y para máquinas de tipo B también están distribuidos normalmente pero con

16. *Fuente:* "SAT escore by Intended Field of Study", Riverside (Calif) Press Enterprise, 8 de abril de 1993.

media μ_2 y varianza $3\sigma^2$. El costo esperado de reparación por semana para la fábrica es entonces $2\mu_1 + \mu_2$. Si tenemos una muestra aleatoria $X_1, X_2, \ldots, X_n$ de los costos de máquinas tipo A y una muestra aleatoria independiente $Y_1, Y_2, \ldots, Y_m$ de los costos para máquinas tipo B, describa cómo construiría un intervalo de confianza de 95% para $2\mu_1 + \mu_2$

a si se conoce σ^2,

b si no se conoce σ^2.

8.94 Suponga que obtenemos muestras independientes de tamaños n_1 y n_2 de dos poblaciones normales con varianzas iguales. Utilice la cantidad pivote apropiada de la Sección 8.8 para deducir un *límite de confianza superior* $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$.

8.9 Intervalos de confianza para σ^2

La varianza poblacional σ^2 cuantifica la cantidad de variabilidad en la población. Muchas veces el valor real de σ^2 es desconocido para un experimentador y debe calcularse. En la Sección 8.3 demostramos que $S^2 = [1/(n-1)] \sum_{i=1}^{n} (Y_i - \overline{Y})^2$ es un estimador insesgado para σ^2. Cuando generamos intervalos de confianza para μ, usamos S^2 para calcular σ^2 cuando era desconocida.

Además de necesitar información acerca de σ^2 para calcular intervalos de confianza para μ y $\mu_1 - \mu_2$, podemos estar interesados en construir un intervalo de confianza para σ^2. Por ejemplo, si efectuamos un cuidadoso análisis químico de tabletas de un medicamento en particular, estaríamos interesados en la cantidad promedio del ingrediente activo por tableta *y además* en la cantidad de variabilidad de una tableta a otra, cuantificada por σ^2. Obviamente, para un medicamento es preferible que la variación de una tableta a otra sea pequeña y por tanto un valor pequeño para σ^2.

Para continuar con nuestro procedimiento de cálculo de intervalos necesitamos la existencia de una cantidad pivote. De nuevo, suponemos que tenemos una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de una distribución normal con media μ y varianza σ^2, ambas desconocidas. Del Teorema 7.3 sabemos que

$$\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución χ^2 con $(n - 1)$ grados de libertad. Entonces podemos proseguir con el método del pivote para hallar dos números χ_L^2 y χ_U^2 tales que

$$P\left[\chi_L^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_U^2\right] = 1 - \alpha$$

para cualquier coeficiente de confianza $(1 - \alpha)$. (Los subíndices L y U representan *bajo* y *alto*, respectivamente.) La función de densidad χ^2 no es simétrica, de modo que tenemos alguna libertad para seleccionar χ_L^2 y χ_U^2. Nos gustaría hallar el intervalo más corto que incluya σ^2 con probabilidad $(1 - \alpha)$. En general, esto es difícil y requiere una búsqueda de prueba y error para los valores apropiados de χ_L^2 y χ_U^2. Elegiremos de manera arbitraria puntos que limiten áreas iguales de cola, como se indica en la Figura 8.11. En consecuencia, obtenemos

$$P\left[\chi_{1-(\alpha/2)}^2 \leq \frac{(n-1)S^2}{\alpha^2} \leq \chi_{(\alpha/2)}^2\right] = 1 - \alpha,$$

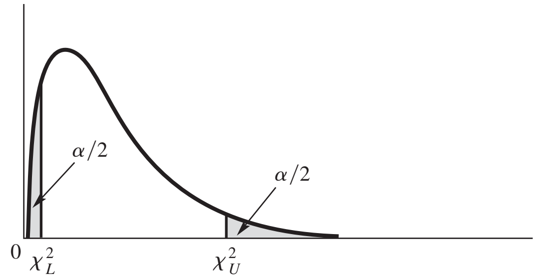


FIGURA **8.11** Ubicación de $\chi^2_{1-(\alpha/2)}$ y $\chi^2_{\alpha/2}$

y un reordenamiento de la desigualdad en el enunciado de probabilidad nos lleva a

$$P\left[\frac{(n-1)S^2}{\chi^2_{(\alpha/2)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right] = 1-\alpha.$$

El intervalo de confianza para σ^2 es el siguiente.

Un intervalo de confianza 100(1 – α)% para σ^2

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}},\ \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right)$$

EJEMPLO **8.13** Un experimentador desea comprobar la variabilidad de mediciones obtenidas al usar equipo diseñado para medir el volumen de una fuente de audio. Tres mediciones independientes registradas por este equipo para la misma fuente de sonido fueron 4.1, 5.2 y 10.2. Estime σ^2 con coeficiente de confianza .90.

Solución Si se puede suponer normalidad en las mediciones registradas por este equipo, se puede aplicar el intervalo de confianza que acabamos de desarrollar. Para los datos dados, $s^2 = 10.57$. Con $\alpha/2 = .05$ y $(n-1) = 2$ grados de libertad, la Tabla 6, Apéndice 3, señala que $\chi^2_{.95} = .103$ y $\chi^2_{.05} = 5.991$. Entonces, el intervalo de confianza de 90% para σ^2 es

$$\left(\frac{(n-1)s^2}{\chi^2_{.05}},\ \frac{(n-1)s^2}{\chi^2_{.95}}\right) \quad \text{o bien} \quad \left(\frac{(2)(10.57)}{5.991},\ \frac{(2)(10.57)}{.103}\right),$$

y finalmente (3.53, 205.24).

Observe que este intervalo para σ^2 es muy ancho, principalmente porque n es muy pequeña. ■

Ya antes hemos indicado que los intervalos de confianza desarrollados en la Sección 8.8 para μ y $\mu_1 - \mu_2$ tenían coeficientes de confianza cercanos al nivel nominal incluso si las poblaciones básicas no estaban distribuidas normalmente. En contraste, los intervalos para σ^2 presentados en esta sección pueden tener coeficientes de confianza que difieren en un modo muy marcado con respecto al nivel nominal si la población muestreada no está distribuida normalmente.

Ejercicios

8.95 La Environmental Protection Agency (EPA) ha establecido un máximo nivel de ruido de 83 decibeles (dB) para camiones pesados. La forma en la que se aplique este límite afectará considerablemente al público y a la industria del transporte por carretera . Una forma de aplicar los límites es exigir que todos los camiones se apeguen al límite de ruido. Un segundo método menos satisfactorio es exigir que el nivel medio de ruido de la flota de camiones sea menor al límite. Si se adopta esta última regla, la variación en el nivel de ruido de un camión a otro se hace importante porque un valor grande de σ^2 implicaría que muchos camiones rebasen ese límite, incluso si el nivel medio de la flota fuera de 83 dB. Una muestra aleatoria de seis camiones pesados produjo los siguientes niveles de ruido (en decibeles):

85.4 86.8 86.1 85.3 84.8 86.0.

Use estos datos para construir un intervalo de confianza de 90% para σ^2, la varianza de las lecturas de emisión de ruido de camiones. Interprete sus resultados.

8.96 En el Ejercicio 8.81 dimos las longitudes de los caparazones de diez langostas maduras *Thenus orientalis* atrapadas en los mares en las cercanías de Singapur. Para su comodidad, los datos se reproducen aquí. Supongamos que usted desea describir la variabilidad de las longitudes del caparazón de esta población de langostas. Encuentre un intervalo de confianza de 90% para la varianza poblacional σ^2.

Número de sector de la langosta	*A061*	*A062*	*A066*	*A070*	*A067*	*A069*	*A064*	*A068*	*A065*	*A063*
Longitud del caparazón (mm)	78	66	65	63	60	60	58	56	52	50

8.97 Suponga que S^2 es la varianza muestral basada en una muestra de tamaño n de una población normal con media y varianza desconocidas. Deduzca un

a límite de confianza superior al $100(1 - \alpha)$ para σ^2

b límite de confianza inferior al $100(1 - \alpha)$ para σ^2.

8.98 Dada una muestra aleatoria de tamaño n de una población normal con media y varianza desconocidas, generamos un intervalo de confianza para la varianza poblacional σ^2 en esta sección. ¿Cuál es la fórmula para un intervalo de confianza para la desviación estándar poblacional σ?

8.99 En el Ejercicio 8.97 dedujimos límites de confianza superior e inferior, cada uno con coeficiente de confianza $1 - \alpha$, para σ^2. ¿Cómo construiríamos un

a límite de confianza superior al $100(1 - \alpha)$ para σ?

b límite de confianza inferior al $100(1 - \alpha)$ para σ?

8.100 Los focos industriales deberían tener una vida media útil aceptable para usuarios potenciales y una variación relativamente pequeña en su duración. Si algunos focos fallan demasiado pronto en su vida útil, los usuarios se molestan y es probable que los cambien por focos producidos por un fabricante diferente. Variaciones grandes por arriba de la media reducen las ventas de reemplazo; en general, la variación en la vida útil de los focos altera los programas de cambio establecidos por los usuarios. Una muestra aleatoria de 20 focos producidos por un fabricante particular produjo los siguientes valores de vida útil (en horas):

2100	2302	1951	2067	2415	1883	2101	2146	2278	2019
1924	2183	2077	2392	2286	2501	1946	2161	2253	1827

Establezca un límite de confianza superior de 99% para la *desviación estándar* de las duraciones de vida útil para los focos producidos por este fabricante. ¿La verdadera desviación estándar poblacional es menor que 150 horas? ¿Por qué sí o por qué no?

8.101 En el trabajo de laboratorio es deseable realizar cuidadosas verificaciones de la variabilidad de lecturas producidas en muestras estándar. En un estudio de la cantidad de calcio en agua potable realizado como parte de una evaluación de calidad del agua, la misma muestra estándar se hizo pasar por el laboratorio seis veces en intervalos aleatorios. Las seis lecturas, en partes por millón, fueron 9.32, 9.48, 9.48, 9.70 y 9.26. Estime la varianza poblacional σ^2 para lecturas en este estándar, usando un intervalo de confianza de 90%.

8.102 Las edades de una muestra aleatoria de cinco profesores universitarios son 39, 54, 61, 72 y 59. Usando esta información encuentre un intervalo de confianza de 99% para la desviación estándar poblacional de las edades de todos los profesores de la universidad, suponiendo que las edades de los profesores universitarios están distribuidas normalmente.

8.103 Un instrumento de precisión está garantizado para dar lecturas que no varían más de 2 unidades. Una muestra de cuatro lecturas del instrumento en el mismo objeto dio las mediciones 353, 351, 351 y 355. Encuentre un intervalo de confianza de 90% para la varianza poblacional. ¿Qué suposiciones son necesarias? ¿Parece razonable la garantía?

8.10 Resumen

El objetivo de muchas investigaciones estadísticas es hacer inferencias acerca de parámetros de la población con base en datos muestrales. Es frecuente que estas inferencias tomen la forma de estimaciones, ya sea puntuales o de intervalo. Preferimos estimadores insesgados con varianza pequeña. La bondad de un estimador insesgado $\hat{\theta}$ puede ser medida por $\sigma_{\hat{\theta}}$ porque el error de estimación es generalmente menor que $2\sigma_{\hat{\theta}}$ con una alta probabilidad. El error cuadrático medio de un estimador, $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$, es pequeño sólo si el estimador tiene varianza pequeña y sesgo pequeño.

Las estimaciones de intervalo de muchos parámetros, por ejemplo μ y p, se pueden obtener a partir de la distribución normal para tamaños muestrales grandes debido al teorema del límite central. Si los tamaños muestrales son pequeños, debe suponerse normalidad de la población y la distribución t se usa para generar intervalos de confianza. No obstante, el intervalo para una sola media es bastante estable en relación con desviaciones moderadas a partir de la normalidad. Esto es, el coeficiente de confianza real asociado con intervalos que tienen un coeficiente nominal de confianza de $100(1 - \alpha)\%$ es muy cercano al nivel nominal incluso si la distribución poblacional difiere moderadamente de la normalidad. El intervalo de confianza para una diferencia en dos medias también es estable en relación con desviaciones moderadas de la normalidad y con respecto a la suposición de varianzas poblacionales iguales si $n_1 \approx n_2$. Cuando n_1 y n_2 se hacen más diferentes, la suposición de varianzas poblacionales iguales se hace más importante.

Si se han seleccionado mediciones muestrales de una distribución normal, se puede desarrollar un intervalo de confianza para σ^2 mediante el uso de la distribución χ^2. Estos intervalos son muy sensibles a la suposición de que la población básica está distribuida normalmente. En consecuencia, el coeficiente de confianza real asociado con el procedimiento de estimación de intervalo puede diferir en forma marcada del valor nominal si la población básica no está distribuida normalmente.

Bibliografía y lecturas adicionales

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

Hoel, P. G. 1984. *Introduction to Mathematical Statistics*, 5th ed. New York: Wiley.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.
Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill.

Ejercicios complementarios

8.104 **Opción múltiple** Se realizó un estudio para determinar qué servicios prefieren los adultos en la telefonía celular. Los resultados del estudio demostraron que 73% de usuarios de teléfonos celulares deseaban servicios de e-mail, con un margen de error de $\pm 4\%$. ¿Qué quiere decir la frase "$\pm 4\%$"?

a Estiman que 4% de la población encuestada puede cambiar de idea entre el tiempo en que se realizó la encuesta y aquel en el que se publicaron los resultados.

b Hay una probabilidad de 4% de que el porcentaje real de usuarios de teléfonos celulares que desean servicio de e-mail no esté en el intervalo (0.69, .077).

c Sólo 4% de la población fue encuestada.

d Sería poco probable obtener la proporción muestral observada de 0.73 a menos que la proporción real de usuarios de teléfonos celulares que desean e-mail esté entre 0.69 y 0.77.

e La probabilidad es .04 de que la proporción muestral esté en el intervalo (0.69, .077).

8.105 Una muestra aleatoria de tamaño 25 se tomó de una población normal con $\sigma^2 = 6$. Un intervalo de confianza para la media se dio como (5.37, 7.37). ¿Cuál es el coeficiente de confianza asociado con este intervalo?

8.106 En un estudio de polinización controlada donde aparece la *Phlox drummondii*, una planta anual que florece en primavera y que es común a lo largo de carreteras en terrenos arenosos en la región central de Texas, Karen Pittman y Donald Levin[17] encontraron que los porcentajes de supervivencia de semillas no eran afectados por la escasez de agua o de nutrientes. En el experimento, las flores en plantas fueron identificadas como machos cuando donaban polen y como hembras cuando eran polinizadas por polen donador en tres grupos de tratamiento: control, con poca agua y con pocos nutrientes. Los datos de la siguiente tabla reflejan un aspecto de los hallazgos del experimento: el número de semillas que sobreviven hasta la madurez de cada uno de los tres grupos para padres machos y hembras.

	Macho		Hembra	
Tratamiento	n	Número de sobrevivientes	n	Número de sobrevivientes
Control	585	543	632	560
Poco agua	578	522	510	466
Pocos nutrientes	568	510	589	546

a Encuentre un intervalo de confianza de 99% para la diferencia entre las proporciones de supervivencia en el grupo de poca agua contra el grupo de pocos nutrientes para padres machos.

b Encuentre un intervalo de confianza de 99% para la diferencia entre las proporciones de supervivencia en pares machos y hembras sometidos a poca agua.

17. *Fuente:* Karen Pittman y Donald Levin, "Effects of Parental Identities and Environment on Components of Crossing Success on *Phlox drummondii*", *American Journal of Botany* 76(3)(1989).

8.107 Consulte el Ejercicio 8.106. Supongamos que usted planea estimar la diferencia en los porcentajes de supervivencia de semillas para padres machos en ambientes de poca agua y pocos nutrientes a no más de .03 con probabilidad .95. Si planea usar un número igual de semillas de padres machos en cada ambiente (es decir, $n_1 = n_2$) ¿qué tan grandes deben ser n_1 y n_2?

8.108 Una química que ha preparado un producto diseñado para matar 60% de un tipo particular de insectos, desea evaluar el porcentaje de muertes causadas por su preparación. ¿Qué tamaño muestral debe usar si desea tener 95% de confianza de que sus resultados experimentales caigan a no más de .02 de la fracción real de insectos muertos?

8.109 Para estimar la proporción de trabajadores desempleados en Panamá, un economista selecciona aleatoriamente 400 personas de la clase trabajadora. De éstas, 25 estaban desempleadas.

a Estime la proporción real de trabajadores desempleados y determine límites para el error de estimación.

b ¿Cuántas personas deben muestrearse para reducir el límite de error de estimación a .02?

8.110 La experiencia pasada demuestra que la desviación estándar del ingreso anual de trabajadores textiles en cierto estado es $400. ¿Cuántos trabajadores textiles sería necesario muestrear si se desea estimar la media poblacional a no más de $50.00, con probabilidad .95?

8.111 ¿Cuántos votantes deben estar incluidos en una muestra recolectada para calcular la fracción del voto favorable a un candidato presidencial en una elección nacional, si la estimación debe ser correcta con tolerancia no mayor que .005? Suponga que la fracción verdadera se encuentra en la cercanía de .5. Use un coeficiente de confianza de aproximadamente .95.

8.112 En una encuesta tomada entre estudiantes universitarios, 300 de entre 500 hombres de la fraternidad estuvieron a favor de cierta proposición, en tanto que 64 de entre 100 que no pertenecían a la fraternidad estaban a favor de la proposición. Estime la diferencia en las proporciones que estaban a favor de la proposición y ponga un límite de desviación estándar de 2 en el error de estimación.

8.113 Consulte el Ejercicio 8.112. ¿Cuántos hombres pertenecientes a la fraternidad y no pertenecientes a ella deben incluirse en una encuesta si deseamos obtener una estimación, con un error máximo de .05, para la diferencia en las proporciones a favor de la proposición? Suponga que los grupos serán de igual tamaño y que $p = .6$ será suficiente como aproximación de ambas proporciones.

8.114 Un proceso químico ha producido, en promedio, 800 toneladas de productos químicos al día. Las producciones diarias de la semana pasada son 785, 805, 790, 793 y 802 toneladas. Estime la producción media diaria, con coeficiente de confianza de .90, a partir de los datos. ¿Qué suposiciones es necesario hacer?

8.115 Consulte el Ejercicio 8.114. Encuentre un intervalo de confianza de 90% para σ^2, la varianza de las producciones diarias.

8.116 ¿Perdemos nuestra capacidad de memoria cuando envejecemos? En un estudio del efecto de la glucosa en la memoria de hombres y mujeres ancianos, C. A. Manning y colegas[18] hicieron una prueba a 16 voluntarios (5 hombres y 11 mujeres) acerca de su memoria de largo plazo, registrando el número de palabras recordadas de una lista leída a cada persona. A cada una de éstas se le recordaron las palabras olvidadas y se le pidió recordar tantas palabras como fuera posible de la lista original. La media y desviación estándar de las calificaciones de memora a largo plazo fueron $\overline{y} = 79.47$ y $s = 25.25$. Proporcione un intervalo de confianza de 99% para las calificaciones de recordar palabras a largo plazo para hombres y mujeres ancianos. Interprete este intervalo.

8.117 El crecimiento anual del tallo principal, medido de una muestra de 17 pinos rojos de 4 años de edad, produjo una media de 11.3 pulgadas y una desviación estándar de 3.4 pulgadas. Encuentre un intervalo de confianza de 90% para el crecimiento anual medio del tallo principal de una población de pinos rojos de 4 años de edad sometidos a condiciones ambientales similares. Suponga que las cantidades de crecimiento están distribuidas normalmente.

18. *Fuente:* C. A. Manning, J. L. Hall y P. E. Gold, "Glucose Effects on Memory and Other Neuropsychological Tests in Elderly Humans", en *Psychological Science* 1(5)(1990).

8.118 Debido a la variabilidad de los descuentos, la utilidad por cada auto nuevo vendido por un distribuidor de autos varía de un auto a otro. Las utilidades por venta (en cientos de dólares), tabuladas para la semana pasada, fueron 2.1, 3.0, 1.2, 6.2, 4.5 y 5.1. Encuentre un intervalo de confianza de 90% para la utilidad media por venta. ¿Qué suposiciones deben ser válidas para que la técnica que usted utilizó sea la apropiada?

8.119 Un examen de matemáticas se aplica a un grupo de 50 estudiantes de la secundaria 1 seleccionados aleatoriamente y también a un grupo de 45 estudiantes de la secundaria 2 seleccionados de la misma manera. Para el grupo de la secundaria 1, la media muestral es 75 puntos y la desviación estándar muestral es 10 puntos. Para el grupo de la secundaria 2, la media muestral es 72 puntos y la desviación estándar muestral es 8 puntos. Construya un intervalo de confianza de 95% para la diferencia en las calificaciones medias. ¿Qué suposiciones son necesarias?

8.120 Dos métodos para enseñar a leer se aplicaron a dos grupos de niños de primaria seleccionados aleatoriamente y se compararon con base en un examen de comprensión de lectura aplicado al final del período de enseñanza. Las medias muestrales y varianzas calculadas a partir de las calificaciones de examen se muestran en la siguiente tabla. Encuentre un intervalo de confianza de 95% para $(\mu_1 - \mu_2)$. ¿Qué suposiciones son necesarias?

Estadístico	Método 1	Método 2
Número de niños en el grupo	11	14
$\overline{y}$	64	69
s^2	52	71

8.121 Una comparación de los tiempos de reacción para dos estímulos diferentes en un experimento psicológico de asociación de palabras produjo los resultados (en segundos) que se muestran en la siguiente tabla cuando se aplicó a una muestra aleatoria de 16 personas. Obtenga un intervalo de confianza de 90% para $(\mu_1 - \mu_2)$. ¿Qué suposiciones son necesarias?

Estímulo 1		Estímulo 2	
1	2	4	1
3	1	2	2
2	3	3	3
1	2	3	3

8.122 El lapso entre la facturación y el pago se registró para una muestra aleatoria de 100 clientes de una empresa de contadores públicos titulados. La media muestral y la desviación estándar para las 100 cuentas fueron 39.1 días y 17.3 días, respectivamente. Encuentre un intervalo de confianza de 90% para la media del tiempo que transcurre entre la facturación y el pago para las 100 cuentas de la empresa de contadores. Interprete el intervalo.

8.123 Los anunciantes en televisión pueden creer erróneamente que casi todas las personas que ven TV entienden la mayor parte de los anuncios que ven y escuchan. Un estudio de investigación reciente pidió a 2300 personas de más de 13 años de edad que vieran extractos de publicidad de televisión de 30 segundos de duración. De éstos, 1914 televidentes entendieron mal todo o parte del extracto que vieron. Encuentre un intervalo de confianza de 95% para la proporción de todos los teleespectadores (de los cuales la muestra es representativa) que entenderán mal el total o parte de los extractos de televisión empleados en este estudio.

8.124 Una encuesta a 415 ejecutivos corporativos, de gobierno y contadores de la Financial Accounting Foundation encontró que 278 consideraban el flujo de caja (lo contrario de ganancias por acción, etc.) como el indicador más importante de la salud financiera de una compañía. Suponga que estos 415 ejecutivos constituyen una muestra aleatoria de la población de todos los ejecutivos. Use los datos para hallar un intervalo de confianza de 95% para la fracción de todos los ejecutivos corporativos que consideran que el flujo de caja es la medida más importante de la salud financiera de una compañía.

8.125 Suponga que muestras independientes de tamaños n_1 y n_2 se toman de dos poblaciones normalmente distribuidas con varianzas σ_1^2 y σ_2^2, respectivamente. Si S_1^2 y S_2^2 denotan las varianzas muestrales respectivas, el Teorema 7.3 implica que $(n_1 - 1)S_1^2/\sigma_1^2$ y $(n_2 - 1)S_2^2/\sigma_2^2$ tienen distribuciones χ^2 con $n_1 - 1$ y $n_2 - 1$ grados de libertad, respectivamente. Además, estas variables aleatorias con distribución χ^2 son independientes porque las muestras se tomaron de manera independiente.

a Utilice estas cantidades para construir una variable aleatoria que tenga una distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

b Utilice la cantidad con distribución F del inciso a como una *cantidad pivote* y deduzca una fórmula para un intervalo de confianza de $100(1 - \alpha)\%$ para σ_2^2/σ_1^2.

8.126 Un fabricante farmacéutico compra materias primas de dos proveedores diferentes. El nivel medio de impurezas es aproximadamente el mismo para ambos proveedores, pero el fabricante está preocupado por la variabilidad en la cantidad de impurezas de entre un embarque y otro. Si el nivel de impurezas tiende a variar en forma excesiva de una fuente de abastecimiento, esto podría afectar la calidad del producto final. Para comparar la variación en el porcentaje de impurezas para los dos proveedores, el fabricante selecciona diez envíos de cada uno de ellos y mide el porcentaje de impurezas de cada envío. Las varianzas muestrales fueron $s_1^2 = .273$ y $s_2^2 = .094$ respectivamente. Forme un intervalo de confianza de 95% para la relación entre las varianzas poblacionales reales.

***8.127** Denote con $\overline{Y}$ la media de una muestra de tamaño 100 tomada de una distribución gamma con $\alpha = c_0$ conocida y β desconocida. Demuestre que un intervalo de confianza aproximado de $100(1 - \alpha)\%$ para β está dado por

$$\left(\frac{\overline{Y}}{c_0 + .1z_{\alpha/2}\sqrt{c_0}}, \frac{\overline{Y}}{c_0 - .1z_{\alpha/2}\sqrt{c_0}}\right).$$

***8.128** Suponga que tomamos una muestra de tamaño n_1 de una población normalmente distribuida con media y varianza μ_1 y σ_1^2, y una muestra independiente de tamaño n_2 de una población normalmente distribuida con media y varianza μ_2 y σ_2^2. Si es razonable suponer que $\sigma_1^2 = \sigma_2^2$, entonces se pueden aplicar los resultados dados en la Sección 8.8.

¿Qué se puede hacer si no podemos suponer que las varianzas desconocidas son iguales, pero tenemos la suerte de saber que $\sigma_2^2 = k\sigma_1^2$ para alguna constante conocida $k \neq 1$? Suponga, como hicimos antes, que las medias muestrales están dadas por $\overline{Y}_1$ y $\overline{Y}_2$ y las varianzas muestrales por S_1^2 y S_2^2, respectivamente.

a Demuestre que Z^* dada a continuación tiene una distribución normal estándar.

$$Z^* = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{\sigma_1\sqrt{\dfrac{1}{n_1} + \dfrac{k}{n_2}}}.$$

b Demuestre que W^* dada a continuación tiene una distribución χ^2 con $n_1 + n_2 - 2$ grados de libertad.

$$W^* = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2/k}{\sigma_1^2}.$$

c Observe que Z^* y W^* de los incisos a y b son independientes. Por último, demuestre que

$$T^* = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{S_p^*\sqrt{\dfrac{1}{n_1} + \dfrac{k}{n_2}}}, \qquad \text{donde } S_p^{2*} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2/k}{n_1 + n_2 - 2}$$

tiene una distribución t con $n_1 + n_2 - 2$ grados de libertad.

d Utilice el resultado del inciso c para dar un intervalo de confianza $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$, suponiendo que $\sigma_2^2 = k\sigma_1^2$.

e ¿Qué ocurre si $k = 1$ en las partes a–d?

***8.129** En la Sección 8.3 observamos que si

$$S'^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n} \quad \text{y} \quad S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1},$$

entonces S^2 es un estimador sesgado de σ^2, pero S^2 es un estimador insesgado del mismo parámetro. Si se toman muestras de una población normal,

a encuentre $V(S^2)$,

b demuestre que $V(S^2) > V(S^2)$.

***8.130** El Ejercicio 8.129 sugiere que S^2 es superior a S'^2 respecto al sesgo y que S'^2 es superior a S^2 porque posee una varianza más pequeña. ¿Cuál es mejor estimador? [*Sugerencia:* compare los errores cuadráticos medios.]

***8.131** Consulte los Ejercicios 1.129 y 1.130. S^2 y S'^2 son dos estimadores para σ^2 de la forma $c\sum_{i=1}^{n}(Y_i - \overline{Y})^2$. ¿Qué valor de c da el estimador para σ^2 con el error cuadrático medio más pequeño entre todas las estimaciones de la forma $c\sum_{i=1}^{n}(Y_i - \overline{Y})^2$?

8.132 Consulte los Ejercicios 6.17 y 8.14. La función de distribución para una distribución de familia de potencias está dada por

$$F(y) = \begin{cases} 0, & y < 0, \\ \left(\dfrac{y}{\theta}\right)^{\alpha}, & 0 \le y \le \theta, \\ 1, & y > \theta, \end{cases}$$

donde $\alpha, \theta > 0$. Suponga que se toma una muestra de tamaño n de una población con una distribución de familia de potencias y que $\alpha = c$ donde $c > 0$ es conocida.

a Demuestre que la función de distribución de $Y_{(n)} = \text{máx}\{Y_1, Y_2, \dots, Y_n\}$ está dada por

$$F_{Y_{(n)}}(y) = \begin{cases} 0, & y < 0, \\ \left(\dfrac{y}{\theta}\right)^{nc}, & 0 \le y \le \theta, \\ 1, & y > \theta, \end{cases}$$

donde $\theta > 0$.

b Demuestre que $Y_{(n)}/\theta$ es una cantidad pivote y que para $0 < k < 1$

$$P\left(k < \frac{Y_{(n)}}{\theta} \le 1\right) = 1 - k^{cn}.$$

c Suponga que $n = 5$ y $\alpha = c = 2.4$.

i Use el resultado del inciso (b) para hallar k tal que

$$P\left(k < \frac{Y_{(5)}}{\theta} \le 1\right) = 0.95.$$

ii Dé un intervalo de confianza de 95% para θ.

***8.133** Suponga que se seleccionan dos muestras aleatorias independientes de n_1 y n_2 observaciones de poblaciones normales. Además, suponga que las poblaciones poseen una varianza común σ^2. Sea

$$S_i^2 = \frac{\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2}{n_i - 1}, \qquad i = 1, 2.$$

a Demuestre que S_p^2, el estimador ponederado de σ^2 (que sigue), es insesgado:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

b Encuentre $V(S_p^2)$.

***8.134** El intervalo de confianza de una muestra pequeña para μ, basado en una t de Student (Sección 8.8) posee una amplitud aleatoria, en contraste con el intervalo de confianza de una muestra grande (Sección 8.6), donde la amplitud no es aleatoria si se conoce σ^2. Encuentre el valor esperado de la amplitud del intervalo en el caso de una muestra pequeña si se desconoce σ^2.

***8.135** Un intervalo de confianza es *insesgado* si el valor esperado del punto medio del intervalo es igual al parámetro estimado. El valor esperado del punto medio del intervalo de confianza de una muestra grande (Sección 8.6) es igual al parámetro estimado y lo mismo es cierto para intervalos de confianza de una muestra pequeña para μ y $(\mu_1 - \mu_2)$ (Sección 8.8). Por ejemplo, el punto medio del intervalo $\bar{y} \pm ts/\sqrt{n}$ es $\bar{y}$, y $E(\overline{Y}) = \mu$. Ahora considere el intervalo de confianza para σ^2. Demuestre que el valor esperado del punto medio de este intervalo de confianza no es igual a σ^2.

***8.136** La media muestral $\overline{Y}$ es un buen estimador puntual de la media poblacional μ. También se puede usar para predecir un valor futuro de Y seleccionado en forma independiente de la población. Supongamos que usted tiene una media muestral $\overline{Y}$ y una varianza S^2 basada en una muestra aleatoria de n mediciones de una población normal. Use la t de Student para formar una cantidad pivote para hallar un intervalo de predicción para algún nuevo valor de Y, por ejemplo Y_p, a ser observado en el futuro. [*Sugerencia:* empiece con la cantidad $\bar{y}_p - \bar{y}$.] Observe la terminología: los parámetros son *estimados;* los valores de variables aleatorias son *pronosticados.*

CAPÍTULO 9

Propiedades de los estimadores puntuales y métodos de estimación

9.1 Introducción

En el Capítulo 8 presentamos algunos estimadores intuitivos para parámetros que con frecuencia son de interés en problemas prácticos. Un estimador $\hat{\theta}$ para un parámetro objetivo θ es una función de las variables aleatorias observadas en una muestra y, por tanto, también es una variable aleatoria. En consecuencia, un estimador tiene una distribución de probabilidad, la *distribución muestral* del estimador. Observamos en la Sección 8.2 que, si $E(\hat{\theta}) = \theta$, entonces el estimador tiene la (a veces) deseable propiedad de ser insesgado.

En este capítulo emprendemos un análisis más formal y detallado de algunas de las propiedades matemáticas de estimadores puntuales, en particular las nociones de eficiencia, consistencia y suficiencia. Presentamos un resultado, el teorema de Rao–Blackwell, que proporciona un enlace entre los estadísticos suficientes y los estimadores insesgados de los parámetros. En términos generales, un estimador insesgado con variación pequeña es, o se puede hacer que sea una función de un estadístico suficiente. También mostraremos un método que en ocasiones puede usarse para hallar estimadores insesgados de mínima varianza para parámetros de interés. A continuación ofrecemos otros dos métodos útiles para obtener estimadores: el método de momentos y el método de máxima verosimilitud. Estudiamos algunas propiedades de estimadores obtenidas por estos métodos.

9.2 Eficiencia relativa

En general es posible obtener más de un estimador insesgado para el mismo parámetro objetivo θ. En la Sección 8.2 (Figura 8.3) mencionamos que si $\hat{\theta}_1$ y $\hat{\theta}_2$ denotan dos estimadores insesgados para el mismo parámetro θ, preferimos usar el estimador con la varianza más pequeña. Esto es, si ambos estimadores son insesgados, $\hat{\theta}_1$ es *relativamente más eficiente* que $\hat{\theta}_2$ si $V(\hat{\theta}_2) > V(\hat{\theta}_1)$. De hecho, usamos la razón $V(\hat{\theta}_2)/V(\hat{\theta}_1)$ para definir la *eficiencia relativa* de dos estimadores insesgados.

DEFINICIÓN **9.1**

Dados dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ de un parámetro θ, con varianzas $V(\hat{\theta}_1)$ y $V(\hat{\theta}_2)$, respectivamente, entonces la *eficiencia* de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$, denotada $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, se define como la razón

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}.$$

Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores insesgados para θ, la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$, $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, es mayor que 1 sólo si $V(\hat{\theta}_2) > V(\hat{\theta}_1)$. En este caso $\hat{\theta}_1$ es un mejor estimador insesgado que $\hat{\theta}_2$. Por ejemplo, si $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = 1.8$, entonces $V(\hat{\theta}_2) = (1.8)V(\hat{\theta}_1)$ y $\hat{\theta}_1$ se prefiere a $\hat{\theta}_2$. Del mismo modo, si $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$ es menor que 1, por ejemplo .73, entonces $V(\hat{\theta}_2) = (.73)V(\hat{\theta}_1)$ y $\hat{\theta}_2$ se prefiere a $\hat{\theta}_1$. Consideremos un ejemplo en donde intervienen dos estimadores diferentes para una media poblacional. Suponga que deseamos estimar la media de una población normal. Sea $\hat{\theta}_1$ la *mediana* muestral, la observación central cuando las mediciones muestrales se ordenan de acuerdo con la magnitud (n impar) o con el promedio de dos observaciones centrales (n par). Sea $\hat{\theta}_2$ la media muestral. Aun cuando se omite la demostración, se puede afirmar que la varianza de la mediana muestral, para n grande, es $V(\hat{\theta}_1) = (1.2533)^2(\sigma^2/n)$. Entonces la eficiencia de la mediana muestral con respecto a la media muestral es

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\sigma^2/n}{(1.2533)^2\sigma^2/n} = \frac{1}{(1.2533)^2} = .6366.$$

Entonces, vemos que la varianza de la media muestral es aproximadamente 64% de la varianza de la mediana muestral. Por tanto, preferiríamos usar la media muestral como el estimador para la media poblacional.

EJEMPLO 9.1 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución uniforme en el intervalo $(0, \theta)$. Dos estimadores insesgados para θ son

$$\hat{\theta}_1 = 2\overline{Y} \text{ y } \hat{\theta}_2 = \left(\frac{n+1}{n}\right) Y_{(n)},$$

cuando $Y_{(n)} = \text{máx}\,(Y_1, Y_2, \ldots, Y_n)$. Encuentre la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$.

Solución En vista que cada Y_i tiene una distribución uniforme en el intervalo $(0, \theta)$, $\mu = E(Y_i) = \theta/2$ y $\sigma^2 = V(Y_i) = \theta^2/12$. Por tanto,

$$E(\hat{\theta}_1) = E(2\overline{Y}) = 2E(\overline{Y}) = 2(\mu) = 2\left(\frac{\theta}{2}\right) = \theta,$$

y $\hat{\theta}_1$ es insesgada, como dijimos. Además,

$$V(\hat{\theta}_1) = V(2\overline{Y}) = 4V(\overline{Y}) = 4\left[\frac{V(Y_i)}{n}\right] = \left(\frac{4}{n}\right)\left(\frac{\theta^2}{12}\right) = \frac{\theta^2}{3n}.$$

Para hallar la media y la varianza de $\hat{\theta}_2$, recuerde (vea Ejercicio 6.74) que la función de densidad de $Y_{(n)}$ está dada por

$$g_{(n)}(y) = n[F_Y(y)]^{n-1} f_Y(y) = \begin{cases} n\left(\dfrac{y}{\theta}\right)^{n-1}\left(\dfrac{1}{\theta}\right), & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Por tanto,

$$E(Y_{(n)}) = \frac{n}{\theta^n}\int_0^\theta y^n\,dy = \left(\frac{n}{n+1}\right)\theta,$$

y se deduce que $E\{[(n+1)/n]Y_{(n)}\} = \theta$; esto es, $\hat{\theta}_2$ es un estimador insesgado para θ.

Como

$$E(Y_{(n)}^2) = \frac{n}{\theta^n}\int_0^\theta y^{n+1}\,dy = \left(\frac{n}{n+2}\right)\theta^2,$$

obtenemos

$$V(Y_{(n)}) = E(Y_{(n)}^2) - [E(Y_{(n)})]^2 = \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2$$

y

$$V(\hat{\theta}_2) = V\left[\left(\frac{n+1}{n}\right)Y_{(n)}\right] = \left(\frac{n+1}{n}\right)^2 V(Y_{(n)})$$
$$= \left[\frac{(n+1)^2}{n(n+2)} - 1\right]\theta^2 = \frac{\theta^2}{n(n+2)}.$$

Por tanto, la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$ está dada por

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\theta^2/[n(n+2)]}{\theta^2/3n} = \frac{3}{n+2}.$$

Esta eficiencia es menor que 1 si $n > 1$. Es decir, si $n > 1$, $\hat{\theta}_2$ tiene una varianza menor que $\hat{\theta}_1$, y por tanto $\hat{\theta}_2$ es en general preferible a $\hat{\theta}_1$ como estimador de θ. ■

Más adelante en este capítulo presentamos algunos métodos para hallar estimadores con varianzas más pequeñas. Por ahora deseamos sólo señalar que la eficiencia relativa es un criterio importante para comparar estimadores.

Ejercicios

9.1 En el Ejercicio 8.8 consideramos una muestra aleatoria de tamaño 3 de una distribución exponencial con función de densidad dada por

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & 0 < y, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

y determinamos que $\hat{\theta}_1 = Y_1, \hat{\theta}_2 = (Y_1 + Y_2)/2, \hat{\theta}_3 = (Y_1 + 2Y_2)/3$ y $\hat{\theta}_5 = \overline{Y}$ son todos ellos estimadores insesgados para θ. Encuentre la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_5$, de $\hat{\theta}_2$ con respecto a $\hat{\theta}_5$ y de $\hat{\theta}_3$ con respecto a $\hat{\theta}_5$.

9.2 Sean $Y_1, Y_2, \ldots, Y_n$ que denotan una muestra aleatoria de una población con media μ y varianza σ^2. Considere los siguientes tres estimadores para μ:

$$\hat{\mu}_1 = \frac{1}{2}(Y_1 + Y_2), \qquad \hat{\mu}_2 = \frac{1}{4}Y_1 + \frac{Y_2 + \cdots + Y_{n-1}}{2(n-2)} + \frac{1}{4}Y_n, \qquad \hat{\mu}_3 = \overline{Y}.$$

a Demuestre que cada uno de los tres estimadores es insesgado.

b Encuentre la eficiencia de $\hat{\mu}_3$ con respecto a $\hat{\mu}_2$ y $\hat{\mu}_1$, respectivamente.

9.3 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución uniforme en el intervalo $(\theta, \theta + 1)$. Sean

$$\hat{\theta}_1 = \overline{Y} - \frac{1}{2} \quad \text{y} \quad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}.$$

a Demuestre que $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores insesgados de θ.

b Encuentre la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$.

9.4 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una distribución uniforme en el intervalo $(0, \theta)$. Si $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$, el resultado del Ejercicio 8.18 indica que $\hat{\theta}_1 = (n+1)Y_{(1)}$ es un estimador insesgado para θ. Si $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$, los resultados del Ejemplo 9.1 implican que $\hat{\theta}_2 = [(n+1)/n]Y_{(n)}$ es otro estimador insesgado para θ. Demuestre que la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$ es $1/n^2$. Observe que esto implica que $\hat{\theta}_2$ es un estimador marcadamente superior.

9.5 Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2. Dos estimadores insesgados de σ^2 son

$$\hat{\sigma}_1^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \quad \text{y} \quad \hat{\sigma}_2^2 = \frac{1}{2}(Y_1 - Y_2)^2.$$

Encuentre la eficiencia de $\hat{\sigma}_1^2$ con respecto a $\hat{\sigma}_2^2$.

9.6 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una distribución de Poisson con media λ. Considere $\hat{\lambda}_1 = (Y_1 + Y_2)/2$ y $\hat{\lambda}_2 = \overline{Y}$. Deduzca la eficiencia de $\hat{\lambda}_1$ con respecto a $\hat{\lambda}_2$.

9.7 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una distribución exponencial con función de densidad dada por

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & 0 < y, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el Ejercicio 8.19 determinamos que $\hat{\theta}_1 = nY_{(1)}$ es un estimador insesgado de θ con $\text{MSE}(\hat{\theta}_1) = \theta^2$. Considere el estimador $\hat{\theta}_2 = \overline{Y}$ y encuentre la eficiencia de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$.

***9.8** Denote con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una función de densidad de probabilidad $f(y)$ que tiene parámetro θ desconocido. Si $\hat{\theta}$ es un estimador insesgado de θ, entonces en condiciones muy generales

$$V(\hat{\theta}) \geq I(\theta), \quad \text{donde } I(\theta) = \left[nE\left(-\frac{\partial^2 \ln f(Y)}{\partial \theta^2}\right)\right]^{-1}.$$

(Esto se conoce como la desigualdad de Cramer–Rao.) Si $V(\hat{\theta}) = I(\theta)$, se dice que el estimador $\hat{\theta}$ es *eficiente.*[1]

a Suponga que $f(y)$ es la densidad normal con media μ y varianza σ^2. Demuestre que $\overline{Y}$ es un estimador eficiente de μ.

b Esta desigualdad también se cumple para funciones de probabilidad discretas $p(y)$. Suponga que $p(y)$ es la función de probabilidad de Poisson con media λ. Demuestre que $\overline{Y}$ es un estimador eficiente de λ.

9.3 Consistencia

Suponga que una moneda, que tiene probabilidad p de resultar en cara, se lanza al aire n veces. Si los tiros son independientes, entonces Y, el número de caras entre los n tiros, tiene una distribución binomial. Si el valor verdadero de p es desconocido, la proporción muestral Y/n es un estimador de p. ¿Qué le ocurre a esta proporción muestral cuando aumenta el número de tiros n? Nuestra intuición nos lleva a pensar que cuando n aumenta, Y/n debe acercarse al verdadero valor de p. Esto es, cuando aumenta la cantidad de información en la muestra, nuestro estimador debe acercarse a la cantidad que valoramos.

La Figura 9.1 ilustra los valores de $\hat{p} = Y/n$ para una sola sucesión de 1000 intentos de Bernoulli cuando el verdadero valor de p es .5. Observe que los valores de $\hat{p}$ convergen alrededor de .5 cuando el número de intentos es pequeño pero se aproximan y permanecen muy cerca de $p = .5$ cuando aumenta el número de intentos.

La sucesión individual de 1000 intentos ilustrada en la Figura 9.1 resultó (para n más grande) en valores para la estimación que fueron muy cercanos al verdadero valor, $p = .5$. ¿Otras sucesiones darían resultados similares? La Figura 9.2 muestra los resultados combinados de 50 sucesiones de 1000 intentos. Observe que las 50 sucesiones distintas no eran idénticas. Más bien, la Figura 9.2 muestra una "convergencia" de clases en el valor real $p = .5$. Esto se ilustra por medio de una dispersión más amplia de los valores de las estimaciones para números de intentos más pequeños, pero una dispersión mucho más angosta de los valores de las estimaciones aparece cuando el número de intentos es más grande. ¿Observaremos este mismo fenómeno para diferentes valores de p? Algunos de los ejercicios que se encuentran al final de esta sección le permitirán usar aplicaciones prácticas (disponibles en www.thomsonedu.com/statistics/wackerly) para explorar más a fondo por sí solo.

¿Cómo podemos expresar técnicamente el tipo de "convergencia" exhibido en la Figura 9.2? Como Y/n es una variable aleatoria, podemos expresar esta "cercanía" a p en términos probabilísticos. En particular, examinemos la probabilidad de que la distancia entre el estimador y el parámetro objetivo, $|(Y/n) - p|$, será menor que algún número real ε positivo arbitra-

1. Los ejercicios precedidos por un asterisco son opcionales.

FIGURA **9.1** Valores de $\hat{p} = Y/n$ para una sucesión individual de 1000 intentos de Bernoulli, $p = .5$

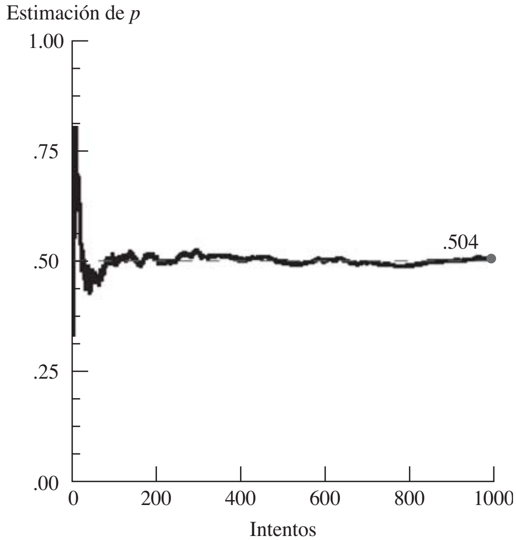
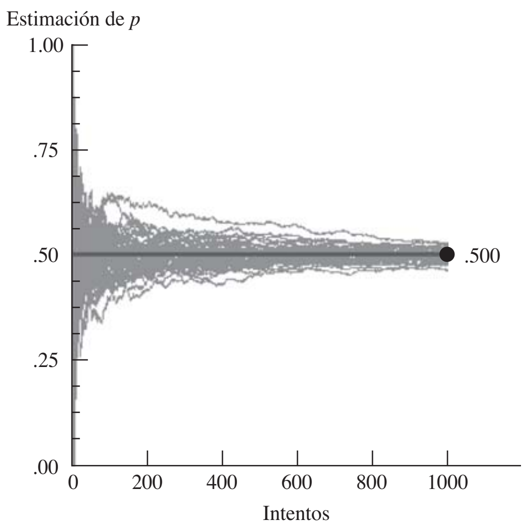


FIGURA **9.2** Valores de $\hat{p} = Y/n$ para 50 sucesiones de 1000 intentos de Bernoulli, $p = .5$



rio. La Figura 9.2 parece indicar que esta probabilidad pudiera ser creciente a medida que n se hace más grande. Si nuestra intuición es correcta y n es grande, esta probabilidad,

$$P\left(\left|\frac{Y}{n} - p\right| \leq \varepsilon\right),$$

debe ser cercana a 1. Si esta probabilidad de hecho tiende a 1 cuando $n \to \infty$, entonces decimos que (Y/n) es un *estimador consistente* de p, o que (Y/n) "converge en probabilidad en p".

DEFINICIÓN **9.2**

Se dice que el estimador $\hat{\theta}_n$ es un *estimador consistente* de θ si, para cualquier número positivo ε,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$$

o bien, de forma equivalente,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

La notación $\hat{\theta}_n$ expresa que el estimador para θ se calcula usando una muestra de tamaño n. Por ejemplo, $\overline{Y}_2$ es el promedio de dos observaciones mientras que $\overline{Y}_{100}$ es el promedio de las 100 observaciones contenidas en una muestra de tamaño $n = 100$. Si $\hat{\theta}_n$ es un estimador insesgado, el siguiente teorema se puede utilizar a menudo para demostrar que el estimador es consistente.

TEOREMA **9.1**

Un estimador insesgado $\hat{\theta}_n$ para θ es un estimador consistente de θ si

$$\lim_{n\to\infty} V(\hat{\theta}_n) = 0.$$

Demostración

Si Y es cualquier variable aleatoria con $E(Y) = \mu$ y $V(Y) = \sigma^2 < \infty$ y si k es cualquier constante no negativa, el teorema de Tchebysheff (véase el Teorema 4.13) implica que

$$P(|Y - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

Como $\hat{\theta}_n$ es un estimador insesgado para θ, se deduce que $E(\hat{\theta}_n) = \theta$. Sea $\sigma_{\hat{\theta}_n} = \sqrt{V(\hat{\theta}_n)}$ que denota el error estándar del estimador $\hat{\theta}_n$. Si aplicamos el teorema de Tchebysheff para la variable aleatoria $\hat{\theta}_n$, obtenemos

$$P\left(\left|\hat{\theta}_n - \theta\right| > k\sigma_{\hat{\theta}_n}\right) \leq \frac{1}{k^2}.$$

Sea n cualquier tamaño muestral fijo. Para cualquier número positivo ε,

$$k = \frac{\varepsilon}{\sigma_{\hat{\theta}_n}}$$

es un número positivo. La aplicación del teorema de Tchebysheff para esta n fija y esta selección de k muestra que

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = P\left(\left|\hat{\theta}_n - \theta\right| > \left[\frac{\varepsilon}{\sigma_{\hat{\theta}_n}}\right]\sigma_{\hat{\theta}_n}\right) \leq \frac{1}{\left(\varepsilon/\sigma_{\hat{\theta}_n}\right)^2} = \frac{V(\hat{\theta}_n)}{\varepsilon^2}.$$

Entonces, para cualquier n fija,

$$0 \leq P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \leq \frac{V(\hat{\theta}_n)}{\varepsilon^2}.$$

Si $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$ y tomamos el límite cuando $n \to \infty$ de la sucesión de probabilidades anterior,

$$\lim_{n\to\infty}(0) \leq \lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \leq \lim_{n\to\infty} \frac{V(\hat{\theta}_n)}{\varepsilon^2} = 0.$$

Entonces, $\hat{\theta}_n$ es un estimador consistente para θ.

La propiedad de consistencia dada en la Definición 9.2 y presentada en el Teorema 9.1 comprende un tipo particular de convergencia de $\hat{\theta}_n$ en θ. Por esta razón, el enunciado "$\hat{\theta}_n$ es un estimador consistente para θ" se sustituye a veces con el enunciado equivalente "$\hat{\theta}_n$ converge en probabilidad en θ."

EJEMPLO **9.2** Sea $Y_1, Y_2, \ldots, Y_n$ que representan una muestra aleatoria de una distribución con media μ y varianza $\sigma^2 < \infty$. Demuestre que $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ es un estimador consistente de μ. (*Nota:* usamos la notación $\overline{Y}_n$ para indicar explícitamente que $\overline{Y}$ se calcula usando una muestra de tamaño n.)

Solución Sabemos de capítulos anteriores que $E(\overline{Y}_n) = \mu$ y $V(\overline{Y}_n) = \sigma^2/n$. Como $\overline{Y}_n$ es insesgado para μ y $V(\overline{Y}_n) \to 0$ cuando $n \to \infty$, el Teorema 9.1 establece que $\overline{Y}_n$ es un estimador consistente de μ. Análogamente, podemos decir que $\overline{Y}_n$ converge en probabilidad en μ.

Al hecho de que $\overline{Y}_n$ sea consistente para μ, o converge en probabilidad en μ, se le conoce a veces como la *ley de los grandes números*. Ésta proporciona la justificación teórica para el proceso de promediar que se emplea en numerosos experimentos para obtener precisión en las mediciones. Por ejemplo, un experimentador puede tomar el promedio de los pesos de muchos animales para obtener una estimación más precisa del promedio de peso de animales de esta especie. La idea del experimentador, confirmada por el Teorema 9.1, es que el promedio de muchos pesos seleccionados de manera independiente debe ser muy cercano al verdadero peso medio con probabilidad alta. ■

En la Sección 8.3 consideramos un estimador intuitivo para $\mu_1 - \mu_2$, la diferencia en las medias de dos poblaciones. El estimador estudiado esa vez fue $\overline{Y}_1 - \overline{Y}_2$, la diferencia en las medias de muestras aleatorias independientes seleccionadas de entre dos poblaciones. Los resultados del Teorema 9.2 serán muy útiles para establecer la consistencia de tales estimadores.

TEOREMA **9.2** Suponga que $\hat{\theta}_n$ converge en probabilidad en θ y que $\hat{\theta}'_n$ converge en probabilidad en θ'.

a $\hat{\theta}_n + \hat{\theta}'_n$ converge en probabilidad en $\theta + \theta'$.
b $\hat{\theta}_n \times \hat{\theta}'_n$ converge en probabilidad en $\theta \times \theta'$.
c Si $\theta' \neq 0$, $\hat{\theta}_n/\hat{\theta}'_n$ converge en probabilidad en θ/θ'.
d Si $g(\cdot)$ es una función de valor real que es continua en θ, entonces $g(\hat{\theta}_n)$ converge en probabilidad en $g(\theta)$.

La demostración del Teorema 9.2 se asemeja mucho a la demostración correspondiente en el caso donde $\{a_n\}$ y $\{b_n\}$ son sucesiones de números reales que convergen en límites reales a y b, respectivamente. Por ejemplo, si $a_n \to a$ y $b_n \to b$ entonces

$$a_n + b_n \to a + b.$$

EJEMPLO **9.3** Suponga que $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria tal que $E(Y_i) = \mu$, $E(Y_i^2) = \mu'_2$ y $E(Y_i^4) = \mu'_4$ son todas finitas. Demuestre que

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

es un estimador consistente de $\sigma^2 = V(Y_i)$. (*Nota:* usamos el subíndice n en S^2 y en $\overline{Y}$ para dar a entender explícitamente su dependencia del valor del tamaño muestral n.)

Solución Hemos visto ya en capítulos anteriores que S^2, que ahora se escribe como S_n^2, es

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}_n^2\right) = \left(\frac{n}{n-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2\right).$$

El estadístico $(1/n)\sum_{i=1}^{n} Y_i^2$ es el promedio de n variables aleatorias independientes y distribuidas idénticamente, con $E(Y_i^2) = \mu'_2$ y $V(Y_i^2) = \mu'_4 - (\mu'_2)^2 < \infty$. Por la ley de los grandes números (Ejemplo 9.2), sabemos que $(1/n)\sum_{i=1}^{n} Y_i^2$ converge en probabilidad a μ'_2.

El Ejemplo 9.2 también implica que $\overline{Y}_n$ converge en probabilidad en μ. Como la función $g(x) = x^2$ es continua para todos los valores finitos de x, el Teorema 9.2(d) implica que $\overline{Y}_n^2$ converge en probabilidad en μ^2. Se deduce entonces del Teorema 9.2(a) que

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2$$

converge en probabilidad en $\mu'_2 - \mu^2 = \sigma^2$. Como $n/(n-1)$ es una sucesión de constantes que convergen en 1 cuando $n \to \infty$, podemos concluir que S_n^2 converge en probabilidad en σ^2. Análogamente, S_n^2, la varianza muestral, es un estimador consistente para σ^2, la varianza poblacional. ■

En la Sección 8.6 consideramos intervalos de confianza de una muestra grande para algunos parámetros de interés práctico. En particular, si $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de cualquier distribución con media μ y varianza σ^2, establecimos que

$$\overline{Y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

es un intervalo de confianza válido de una muestra grande con coeficiente de confianza aproximadamente igual a $(1 - \alpha)$. Si σ^2 se conoce, este intervalo puede y debe calcularse. No obstante, si σ^2 no se conoce pero el tamaño muestral es grande, recomendamos sustituir S por σ en el cálculo porque esto no ocasiona una pérdida significativa de exactitud. El siguiente teorema proporciona la justificación teórica para estas afirmaciones.

TEOREMA 9.3

Suponga que U_n tiene una función de distribución que converge en una función de distribución normal estándar cuando $n \to \infty$. Si W_n converge en probabilidad en 1, entonces la función de distribución de U_n/W_n converge en una función de distribución normal estándar.

Este resultado se deduce de un resultado general conocido como *teorema de Slutsky* (Serfling, 2002). La demostración de este resultado está fuera del propósito de este libro, pero la utilidad del resultado se ilustra en el siguiente ejemplo.

EJEMPLO 9.4 Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de tamaño n de una distribución con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2$. Defina S_n^2 como

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2.$$

Demuestre que la función de distribución de

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converge en una función de distribución normal estándar.

Solución En el Ejemplo 9.3, demostramos que S_n^2 converge en probabilidad en σ^2. Observe que $g(x) = +\sqrt{x/c}$ es una función continua de x si x y c son positivas. Por tanto, se deduce del Teorema 9.2(d) que $S_n/\sigma = +\sqrt{S_n^2/\sigma^2}$ converge en probabilidad en 1. También sabemos del teorema del límite central (Teorema 7.4) que la función de distribución de

$$U_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right)$$

converge en una función de distribución normal estándar. Por tanto, el Teorema 9.3 implica que la función de distribución de

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right)\Big/(S_n/\sigma) = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converge en una función de distribución normal estándar. ■

El resultado del Ejemplo 9.4 indica que, cuando n es grande, $\sqrt{n}(\overline{Y}_n - \mu)/S_n$ tiene aproximadamente una distribución normal estándar *cualquiera que sea* la forma de la distribución de la que se tome la muestra. Si la muestra se toma de una *distribución normal,* los resultados del Capítulo 7 implican que $t = \sqrt{n}(\overline{Y}_n - \mu)/S_n$ tiene una distribución t con $n - 1$ grados de libertad (gl). Combinando esta información, vemos que si una muestra grande se toma de una distribución normal, la función de distribución de $t = \sqrt{n}(\overline{Y}_n - \mu)/S_n$ puede ser aproximada por una función de distribución normal estándar. Esto es, cuando n se hace grande y por tanto cuando el número de grados de libertad es grande, la función de distribución t converge en la función de distribución normal estándar.

Si obtenemos una muestra grande de cualquier distribución, sabemos del Ejemplo 9.4 que $\sqrt{n}(\overline{Y}_n - \mu)/S_n$ tiene aproximadamente una distribución normal estándar. Por tanto, se deduce que

$$P\left[-z_{\alpha/2} \leq \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right) \leq z_{\alpha/2}\right] \approx 1 - \alpha.$$

Si manipulamos las desigualdades del enunciado de probabilidad para aislar μ en el centro, obtenemos

$$P\left[\overline{Y}_n - z_{\alpha/2}\left(\frac{S_n}{\sqrt{n}}\right) \leq \mu \leq \overline{Y}_n + z_{\alpha/2}\left(\frac{S_n}{\sqrt{n}}\right)\right] \approx 1 - \alpha.$$

Por tanto, $\overline{Y}_n \pm z_{\alpha/2}(S_n/\sqrt{n})$ forma un intervalo de confianza de muestra grande válido para μ, con coeficiente de confianza aproximadamente igual a $1 - \alpha$. Del mismo modo, el Teorema 9.3 se puede aplicar para demostrar que

$$\hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_n\hat{q}_n}{n}}$$

es un intervalo de confianza válido con una muestra grande para p con coeficiente de confianza aproximadamente igual a $1 - \alpha$.

En esta sección hemos visto que la propiedad de consistencia es un indicador de la distancia entre un estimador y la cantidad que se estima. Hemos visto que cuando el tamaño muestral es grande, $\overline{Y}_n$ es cercana a μ y S_n^2 es cercana a σ^2, con una probabilidad alta. Más adelante en este capítulo veremos otros ejemplos de estimadores consistentes.

En esta sección utilizamos la notación $\overline{Y}_n$, S_n^2, $\hat{p}_n$ y, en general, $\hat{\theta}_n$ para dar a entender explícitamente la dependencia de los estimadores respecto del tamaño muestral n. Necesitábamos hacerlo así porque estábamos interesados en calcular

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon).$$

Si este límite es igual a 1, entonces $\hat{\theta}_n$ es un estimador "consistente" para θ (más precisamente, $\hat{\theta}_n$ es una *sucesión de estimadores* consistente para θ). Desafortunadamente, esta notación hace que nuestros estimadores se vean demasiado complejos. De aquí en adelante, volveremos a utilizar la notación $\hat{\theta}$ como nuestro estimador para θ y no mostraremos explícitamente la dependencia del estimador en n. La dependencia de $\hat{\theta}$ respecto del tamaño muestral n siempre es implícita y debe tomarse en cuenta cuando se considere la consistencia del estimador.

Ejercicios

9.9 **Ejercicio Applet** ¿Cómo se obtuvo la Figura 9.1? Entre a la aplicación breve *PointSingle* en www.thomsonedu.com/statistics/wackerly. La aplicación superior va a generar una sucesión de intentos de Bernoulli [$X_i = 1, 0$ con $p(1) = p, p(0) = 1 - p$] con $p = .5$, un escenario equivalente a sucesivamente lanzar al aire una moneda balanceada. Sea $Y_n = \sum_{i=1}^{n} X_i$ = el número de unos (1) en los primeros n intentos y $\hat{p}_n = Y_n/n$. Para cada n, la aplicación calcula p_n y la grafica contra el valor de n.

a Si $\hat{p}_5 = 2/5$, ¿qué valor de X_6 resultará en $\hat{p}_6 > \hat{p}_5$?

b Haga clic en el botón "One Trial" una sola vez. Su primera observación es 0 o 1. ¿Qué valor obtuvo? ¿Cuál fue el valor de $\hat{p}_1$? Haga clic en el botón "One Trial" varias veces más. ¿Cuántos intentos n ha simulado? ¿Qué valor de $\hat{p}_n$ observó? ¿El valor es cercano a .5, el valor real de p? ¿La gráfica es una línea horizontal plana? ¿Por qué sí o por qué no?

c Haga clic en el botón "100 Trials" una sola vez. ¿Qué observa? Haga clic en el botón "100 Trials" repetidas veces hasta que el número total de intentos sea 1000. ¿La gráfica que obtuvo es idéntica a la dada en la Figura 9.1? ¿En qué sentido es semejante a la gráfica de la Figura 9.1?

d Con base en la muestra de tamaño 1000, ¿cuál es el valor de $\hat{p}_{1000}$? ¿Este valor es lo que esperaba observar?

e Haga clic en el botón "Reset". Haga clic en el botón "100 Trials" diez veces para generar otra sucesión de valores para $\hat{p}$. Coméntelo.

9.10 **Ejercicio Applet** Consulte el Ejercicio 9.9. Arrastre el cursor a la parte de la pantalla marcada "Try different probabilities". Use el botón marcado "p =" de la esquina inferior derecha de la pantalla para cambiar el valor de p a un valor que no sea .5.

a Haga clic en el botón "One Trial" unas cuantas veces. ¿Qué observa?

b Haga clic en el botón "100 Trials" unas cuantas veces. ¿Qué observa acerca de los valores de $\hat{p}_n$ cuando el número de intentos se hace más grande?

9.11 **Ejercicio Applet** Consulte los Ejercicios 9.9 y 9.10. ¿Cómo se pueden graficar simultáneamente los resultados de varias sucesiones de intentos de Bernoulli? Entre a la aplicación *PointbyPoint.* Arrastre el cursor hasta que pueda ver los seis botones bajo la gráfica superior.

a No cambie el valor preestablecido de $p = .5$. Haga clic en el botón "One Trial" unas cuantas veces para verificar que esté obteniendo un resultado semejante a los obtenidos en el Ejercicio 9.9. Haga clic en el botón "5 Trials" hasta que haya generado un total de 50 intentos. ¿Cuál es el valor de $\hat{p}_{50}$ que obtuvo al final de esta primera sucesión de 50 intentos?

b Haga clic en el botón "New Sequence". El color de su gráfica inicial cambia de rojo a verde. Haga clic en el botón "5 Trials" unas cuantas veces. ¿Qué observa? ¿La gráfica es igual a la que observó en el inciso a? ¿En qué sentido es similar?

c Haga clic en el botón "New Sequence". Genere una nueva sucesión de 50 intentos. Repita hasta que haya generado cinco sucesiones. ¿Las trayectorias generadas por las cinco sucesiones son idénticas? ¿En qué sentido son similares?

9.12 **Ejercicio Applet** Consulte el Ejercicio 9.11. ¿Qué ocurre si cada sucesión es más larga? Arrastre el cursor a la parte de la porción marcada "Longer Sequences of Trials".

a Repita las instrucciones en los incisos a–c del Ejercicio 9.11.

b ¿Qué espera que ocurra si p no es 0.5? Use el botón de la esquina inferior derecha para cambiar el valor de p. Genere varias sucesiones de intentos. Coméntelo.

9.13 **Ejercicio Applet** Consulte los Ejercicios 9.9–9.12. Entre a la aplicación *Point Estimation.*

a Seleccione un valor para p. Haga clic en el botón "New Sequence" repetidas veces. ¿Qué observa?

b Arrastre el cursor a la parte de la aplicación marcada "More Trials". Seleccione un valor para p y haga clic en el botón "New Sequence" repetidas veces. Obtendrá hasta 50 sucesiones, cada una basada en 1000 intentos. ¿Cómo cambia la variabilidad entre las estimaciones en función del tamaño muestral? ¿Cómo se manifiesta esto en la pantalla que obtuvo?

9.14 **Ejercicio Applet** Consulte el Ejercicio 9.13. Arrastre el cursor a la parte de la aplicación marcada "Mean of Normal Data". Valores sucesivos observados de una variable aleatoria normal estándar pueden generarse y emplearse para calcular el valor de la media muestral $\overline{Y}_n$. Estos valores sucesivos se grafican entonces contra el tamaño muestral respectivo para obtener una "trayectoria muestral".

a ¿Espera usted que los valores de $\overline{Y}_n$ se agrupen alrededor de algún valor en particular? ¿Qué valor?

b Si se grafican los resultados de 50 trayectorias muestrales, ¿cómo espera que cambie la variabilidad de las estimaciones en función del tamaño muestral?

c Haga clic en el botón "New Sequence" varias veces. ¿Observó lo que esperaba con base en sus respuestas a los incisos a y b?

9.15 Consulte el Ejercicio 9.3. Demuestre que $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores consistentes para θ.

9.16 Consulte el Ejercicio 9.5. ¿$\hat{\sigma}_2^2$ es un estimador consistente de σ^2?

9.17 Suponga que $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_n$ son muestras aleatorias independientes provenientes de poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Demuestre que $\overline{X} - \overline{Y}$ es un estimador consistente de $\mu_1 - \mu_2$.

9.18 En el Ejercicio 9.17 suponga que las poblaciones están distribuidas normalmente con $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Demuestre que

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(Y_i - \overline{Y})^2}{2n - 2}$$

es un estimador consistente de σ^2.

9.19 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\theta > 0$. Demuestre que $\overline{Y}$ es un estimador consistente de $\theta/(\theta + 1)$.

9.20 Si Y tiene una distribución binomial con n intentos y probabilidad de éxito p, demuestre que Y/n es un estimador consistente de p.

9.21 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una población normal con media μ y varianza σ^2. Suponiendo que $n = 2k$ para algún entero k, un posible estimador para σ^2 está dado por

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a Demuestre que $\hat{\sigma}^2$ es un estimador insesgado para σ^2.

b Demuestre que $\hat{\sigma}^2$ es un estimador consistente para σ^2.

9.22 Consulte el Ejercicio 9.21. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de tamaño n de una población con distribución de Poisson con media λ. De nuevo, suponga que $n = 2k$ para algún entero k. Considere

$$\hat{\lambda} = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a Demuestre que $\hat{\lambda}$ es un estimador insesgado para λ.

b Demuestre que $\hat{\lambda}$ es un estimador consistente para λ.

9.23 Consulte el Ejercicio 9.21. Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de tamaño n de una población para la cual los primeros cuatro momentos son finitos. Esto es, $m_1' = E(Y_1) < \infty$, $m_2' = E(Y_1^2) < \infty$, $m_3' = E(Y_1^3) < \infty$ y $m_4' = E(Y_1^4) < \infty$. (*Nota:* esta suposición es válida para las distribuciones normal y de Poisson en los Ejercicios 9.21 y 9.22, respectivamente.) De nuevo, suponga que

$n = 2k$ para algún entero k. Considere

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a Demuestre que $\hat{\sigma}^2$ es un estimador insesgado para σ^2.

b Demuestre que $\hat{\sigma}^2$ es un estimador consistente para σ^2.

c ¿Por qué fue necesaria la suposición de que $m'_4 = E(Y_1^4) < \infty$?

9.24 Sean $Y_1, Y_2, Y_3, \ldots, Y_n$ variables aleatorias normales estándar independientes.

a ¿Cuál es la distribución de $\sum_{i=1}^{n} Y_i^2$?

b Sea $W_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$. ¿$W_n$ *converge en probabilidad* en alguna constante? Si es así, ¿cuál es el valor de la constante?

9.25 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una distribución normal con media μ y varianza 1. Considere la primera observación Y_1 como un estimador para μ.

a Demuestre que Y_1 es un estimador insesgado para μ.

b Encuentre $P(|Y_1 - \mu| \leq 1)$.

c Vea la definición básica de consistencia dada en la Definición 9.2. Con base en el resultado del inciso b, ¿Y_1 es un estimador consistente para μ?

***9.26** En ocasiones es relativamente fácil establecer consistencia o falta de consistencia si se apela directamente a la Definición 9.2, evaluando $P(|\hat{\theta}_n - \theta| \leq \varepsilon)$ de manera directa y luego demostrando que $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$. Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una distribución uniforme en el intervalo $(0, \theta)$. Si $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$, demostramos en el Ejercicio 6.74 que la función de distribución de probabilidad de $Y_{(n)}$ está dada por

$$F_{(n)}(y) = \begin{cases} 0, & y < 0, \\ (y/\theta)^n, & 0 \leq y \leq \theta, \\ 1, & y > \theta. \end{cases}$$

a Para cada $n \geq 1$ y toda $\varepsilon > 0$, se deduce que $P(|Y_{(n)} - \theta| \leq \varepsilon) = P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon)$. Si $\varepsilon > \theta$, verifique que $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1$ y que, para toda $\varepsilon < \theta$ positiva, obtenemos $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1 - [(\theta - \varepsilon)/\theta]^n$.

b Usando el resultado del inciso a, demuestre que $Y_{(n)}$ es un estimador consistente para θ al demostrar que, para toda $\varepsilon > 0$, $\lim_{n\to\infty} P(|Y_{(n)} - \theta| \leq \varepsilon) = 1$.

***9.27** Use el método descrito en el Ejercicio 9.26 para demostrar que, si $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ cuando $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias uniformes e independientes en el intervalo $(0, \theta)$, entonces $Y_{(1)}$ *no* es un estimador consistente para θ. [*Sugerencia:* con base en los métodos de la Sección 6.7, $Y_{(1)}$ tiene la función de distribución

$$F_{(1)}(y) = \begin{cases} 0, & y < 0, \\ 1 - (1 - y/\theta)^n, & 0 \leq y \leq \theta, \\ 1, & y > \theta.] \end{cases}$$

***9.28** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una distribución de Pareto (véase el Ejercicio 6.18). Entonces los métodos de la Sección 6.7 implican que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ tiene la función de distribución dada por

$$F_{(1)}(y) = \begin{cases} 0, & y \leq \beta, \\ 1 - (\beta/y)^{\alpha n}, & y > \beta. \end{cases}$$

Use el método descrito en el Ejercicio 9.26 para demostrar que $Y_{(1)}$ es un estimador consistente de β.

***9.29** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una distribución de familia de potencias (véase el Ejercicio 6.17). Entonces los métodos de la Sección 6.7 implican que $Y_{(n)} = \text{máx}\,(Y_1, Y_2, \ldots, Y_n)$ tiene la función de distribución dada por

$$F_{(n)}(y) = \begin{cases} 0, & y < 0, \\ (y/\theta)^{\alpha n}, & 0 \le y \le \theta, \\ 1, & y > \theta. \end{cases}$$

Use el método descrito en el Ejercicio 9.26 para demostrar que $Y_{(n)}$ es un estimador consistente de θ.

9.30 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes, cada una con función de densidad de probabilidad

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $\overline{Y}$ converge en probabilidad en alguna constante y determine la constante.

9.31 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución gamma con parámetros α y β, demuestre que $\overline{Y}$ converge en probabilidad en alguna constante y determine la constante.

9.32 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y) = \begin{cases} \dfrac{2}{y^2}, & y \ge 2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

¿La ley de grandes números aplica a $\overline{Y}$ en este caso? ¿Por qué sí o por qué no?

9.33 Un experimentador desea comparar los números de bacterias de tipos A y B en muestras de agua. Se toma un total de n muestras de agua independientes y se hacen las cuentas para cada muestra. Sea X_i el número de bacterias tipo A y Y_i el número de bacterias tipo B para la muestra i. Suponga que los dos tipos de bacterias están distribuidos escasamente dentro de una muestra de agua, de modo que $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_n$ se pueden considerar muestras aleatorias independientes de distribuciones de Poisson con medias λ_1 y λ_2, respectivamente. Sugiera un estimador de $\lambda_1/(\lambda_1 + \lambda_2)$. ¿Qué propiedades tiene el estimador propuesto por usted?

9.34 La función de densidad de Rayleigh está dada por

$$f(y) = \begin{cases} \left(\dfrac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el Ejercicio 6.34(a), usted estableció que Y^2 tiene una distribución exponencial con media θ. Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución de Rayleigh, demuestre que $W_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$ es un estimador consistente para θ.

9.35 Sean $Y_1, Y_2, \ldots$ una sucesión de variables aleatorias con $E(Y_i) = \mu$ y $V(Y_i) = \sigma_i^2$. Observe que las σ_i^2 no son todas iguales.

a ¿Cuál es $E(\overline{Y}_n)$?

b ¿Cuál es $V(\overline{Y}_n)$?

c ¿En qué condición (en las σ_i^2) puede aplicarse el Teorema 9.1 para demostrar que $\overline{Y}_n$ es un estimador consistente para μ?

9.36 Suponga que Y tiene una distribución binomial con base en n intentos y probabilidad de éxito p. Entonces $\hat{p}_n = Y/n$ es un estimador insesgado de p. Use el Teorema 9.3 para demostrar que la distribución de

$(\hat{p}_n - p)/\sqrt{\hat{p}_n\hat{q}_n/n}$ converge en una distribución normal estándar. [*Sugerencia:* escriba Y como lo hicimos en la Sección 7.5.]

9.4 Suficiencia

Hasta aquí hemos seleccionado estimadores con base en la intuición. Por tanto, elegimos $\overline{Y}$ y S^2 como los estimadores de la media y la varianza, respectivamente, de la distribución normal. (*Parece* que éstos deben ser buenos estimadores de los parámetros poblacionales.) Hemos visto que en ocasiones es más conveniente usar estimadores insesgados. De hecho, se ha demostrado que $\overline{Y}$ y S^2 son estimadores insesgados de la media poblacional μ y la varianza σ^2, respectivamente. Observe que hemos empleado la información en una muestra de tamaño n para calcular el valor de dos estadísticos que funcionan como estimadores para los parámetros de interés. En esta etapa los valores muestrales reales ya no son importantes; más bien, resumimos la información de la muestra que se relaciona con los parámetros de interés al usar los estadísticos $\overline{Y}$ y S^2. ¿Este proceso de resumir o reducir los datos a los dos estadísticos $\overline{Y}$ y S^2 conserva toda la información acerca de μ y σ^2 en el conjunto original de n observaciones muestrales? O bien, ¿se ha perdido u ocultado alguna información acerca de estos parámetros en el proceso de reducir los datos? En esta sección presentamos métodos para hallar estadísticos que en cierto sentido resumen *toda* la información de una muestra acerca de un parámetro objetivo. Se dice que estos estadísticos tienen la propiedad de *suficiencia* o, dicho en una forma más sencilla, reciben el nombre de *estadísticos suficientes.* Como veremos en la siguiente sección, "buenos" estimadores son (o se puede hacer que sean) funciones de cualquier estadístico suficiente. En realidad, los estadísticos suficientes a menudo se pueden usar para desarrollar estimadores que tienen varianza mínima entre todos los estimadores insesgados.

Para ilustrar la noción de un estadístico suficiente, consideremos los resultados de n intentos de un experimento binomial, $X_1, X_2, \ldots, X_n$, donde

$$X_i = \begin{cases} 1, & \text{si el } i\text{-ésimo intento es un éxito,} \\ 0, & \text{si el } i\text{-ésimo intento es un fracaso.} \end{cases}$$

Si p es la probabilidad de éxito en cualquier intento, entonces, para $i = 1, 2, \ldots, n$,

$$X_i = \begin{cases} 1, & \text{con probabilidad } p, \\ 0, & \text{con probabilidad } q = 1 - p. \end{cases}$$

Suponga que nos dan un valor de $Y = \sum_{i=1}^{n} X_i$, el número de éxitos entre los n intentos. Si conocemos el valor de Y, ¿podemos obtener alguna información adicional acerca de p al ver otras funciones de $X_1, X_2, \ldots, X_n$? Una forma de responder esta pregunta es ver la distribución condicional de $X_1, X_2, \ldots, X_n$, dada Y:

$$P(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, Y = y)}{P(Y = y)}.$$

El numerador del lado derecho de esta expresión es 0 si $\sum_{i=1}^{n} x_i \neq y$, y es la probabilidad de una sucesión independiente de números 0 y 1 con un total de y números 1 y $(n - y)$ números 0 si $\sum_{i=1}^{n} x_i = y$. De la misma manera, el denominador es la probabilidad binomial de exactamente y éxitos en n intentos. Por tanto, si $y = 0, 1, 2, \ldots, n$,

$$P(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \begin{cases} \dfrac{p^y(1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \dfrac{1}{\binom{n}{y}}, & \text{si } \displaystyle\sum_{i=1}^{n} x_i = y, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Es importante observar que la distribución de $X_1, X_2, \ldots, X_n$, dada Y, *no* depende de p. Esto es, una vez que se conozca Y, ninguna otra función de $X_1, X_2, \ldots, X_n$ proporcionará más información sobre el posible valor de p. En este sentido, Y contiene toda la información acerca de p. Por tanto, se dice que el estadístico Y es *suficiente* para p. Generalizamos esta idea en la definición siguiente.

DEFINICIÓN 9.3

Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de probabilidad con parámetro desconocido θ. Entonces se dice que el estadístico $U = g(Y_1, Y_2, \ldots, Y_n)$ es *suficiente* para θ si la distribución condicional de $Y_1, Y_2, \ldots, Y_n$, dada U, no depende de θ.

En muchas exposiciones anteriores hemos considerado que la función de probabilidad $p(y)$, asociada con una variable aleatoria discreta [o la función de densidad $f(y)$ para una variable aleatoria continua], son funciones sólo del argumento y. Nuestras futuras exposiciones se simplifican si adoptamos una notación que nos permita explícitamente mostrar el hecho de que la distribución, asociada con una variable aleatoria Y, depende a menudo del valor de un parámetro θ. Si Y es una variable aleatoria discreta que tiene una función de masa de probabilidad que depende del valor de un parámetro θ, en lugar de $p(y)$ usamos la notación $p(y|\theta)$. Del mismo modo, indicaremos la dependencia explícita de la forma de una función de densidad continua respecto del valor de un parámetro θ si escribimos la función de densidad como $f(y|\theta)$ en lugar de la $f(y)$ empleada antes.

La Definición 9.3 nos dice cómo comprobar si un estadístico es suficiente, pero no nos dice cómo *hallar* un estadístico suficiente. Recuerde que, en el caso discreto, la distribución conjunta de variables aleatorias discretas $Y_1, Y_2, \ldots, Y_n$ está dada por una función de probabilidad $p(y_1, y_2, \ldots, y_n)$. Si esta función de probabilidad conjunta depende explícitamente del valor de un parámetro θ, la escribimos como $p(y_1, y_2, \ldots, y_n | \theta)$. Esta función da la probabilidad o *verosimilitud* de observar el evento $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ cuando el valor del parámetro es θ. En el caso continuo, cuando la distribución conjunta de $Y_1, Y_2, \ldots, Y_n$ depende del parámetro θ, escribiremos la función de densidad conjunta como $f(y_1, y_2, \ldots, y_n | \theta)$. De aquí en adelante, será conveniente tener un solo nombre para la función que defina la distribución conjunta de las variables $Y_1, Y_2, \ldots, Y_n$ observadas en una muestra.

DEFINICIÓN 9.4

Sean $y_1, y_2, \ldots, y_n$ observaciones muestrales tomadas de variables aleatorias correspondientes $Y_1, Y_2, \ldots, Y_n$ cuya distribución depende de un parámetro θ. Entonces, si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias discretas, la *verosimilitud de la muestra,* $L(y_1, y_2, \ldots, y_n | \theta)$, se define como la probabilidad conjunta de $y_1, y_2, \ldots, y_n$.

> Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias continuas, la verosimilitud $L(y_1, y_2, \ldots, y_n \mid \theta)$ se define como la densidad conjunta evaluada en $y_1, y_2, \ldots, y_n$.

Si el conjunto de variables aleatorias $Y_1, Y_2, \ldots, Y_n$ denota una muestra aleatoria de una distribución discreta con función de probabilidad $p(y \mid \theta)$, entonces

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \theta) &= p(y_1, y_2, \ldots, y_n \mid \theta) \\ &= p(y_1 \mid \theta) \times p(y_2 \mid \theta) \times \cdots \times p(y_n \mid \theta), \end{aligned}$$

mientras que si $Y_1, Y_2, \ldots, Y_n$ tienen una distribución continua con función de densidad $f(y \mid \theta)$, entonces

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) \\ &= f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta). \end{aligned}$$

Para simplificar la notación, a veces expresaremos la verosimilitud como $L(\theta)$ en lugar de $L(y_1, y_2, \ldots, y_n \mid \theta)$.

El siguiente teorema relaciona la propiedad de suficiencia con la verosimilitud $L(\theta)$.

TEOREMA 9.4

> Sea U un estadístico basado en la muestra aleatoria $Y_1, Y_2, \ldots, Y_n$. Entonces U es un *estadístico suficiente* para la estimación de un parámetro θ si y sólo si la verosimilitud $L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta)$ se puede factorizar en dos funciones no negativas,
>
> $$L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta) \times h(y_1, y_2, \ldots, y_n$$
>
> donde $g(u, \theta)$ es una función sólo de u y θ y $h(y_1, y_2, \ldots, y_n)$ no es una función de θ.

Aun cuando la demostración del Teorema 9.4 (también conocido como el *criterio de factorización*) está fuera del propósito de este libro, ilustramos la utilidad del teorema en el siguiente ejemplo.

EJEMPLO 9.5 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria en la que Y_i posee la función de densidad de probabilidad

$$f(y_i \mid \theta) = \begin{cases} (1/\theta)e^{-y_i/\theta}, & 0 \leq y_i < \infty, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\theta > 0$, $i = 1, 2, \ldots, n$. Demuestre que $\overline{Y}$ es un estadístico suficiente para el parámetro θ.

Solución La verosimilitud $L(\theta)$ de la muestra es la densidad conjunta

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) \\ &= f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta) \\ &= \frac{e^{-y_1/\theta}}{\theta} \times \frac{e^{-y_2/\theta}}{\theta} \times \cdots \times \frac{e^{-y_n/\theta}}{\theta} = \frac{e^{-\sum y_i/\theta}}{\theta^n} = \frac{e^{-n\overline{y}/\theta}}{\theta^n}. \end{aligned}$$

Observe que $L(\theta)$ es una función sólo de θ y $\overline{y}$ y que si

$$g(\overline{y}, \theta) = \frac{e^{-n\overline{y}/\theta}}{\theta^n} \quad \text{y} \quad h(y_1, y_2, \ldots, y_n) = 1,$$

entonces

$$L(y_1, y_2, \ldots, y_n \mid \theta) = g(\overline{y}, \theta) \times h(y_1, y_2, \ldots, y_n)$$

En consecuencia, el Teorema 9.4 implica que $\overline{Y}$ es un estadístico suficiente para el parámetro θ. ■

El Teorema 9.4 se puede usar para demostrar que hay muchos posibles estadísticos suficientes para cualquier parámetro poblacional. Primero que nada, de acuerdo con la Definición 9.3 o el criterio de factorización (Teorema 9.4), la muestra aleatoria por sí misma es un estadístico suficiente. En segundo término, si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución con una función de densidad con parámetro θ, entonces el conjunto de estadísticos de orden $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, que es una función de $Y_1, Y_2, \ldots, Y_n$, es suficiente para θ. En el Ejemplo 9.5 decidimos que $\overline{Y}$ es un estadístico suficiente para la estimación de θ. El Teorema 9.4 también podría haberse usado para demostrar que $\sum_{i=1}^{n} Y_i$ es otro estadístico suficiente. De hecho, para la distribución exponencial descrita en el Ejemplo 9.5, cualquier estadístico que sea una función biunívoca de $\overline{Y}$ es un estadístico suficiente.

En nuestro ejemplo inicial de esta sección, que comprende el número de éxitos en n intentos, $Y = \sum_{i=1}^{n} X_i$ reduce los datos $X_1, X_2, \ldots, X_n$ a un solo valor que permanece suficiente para p. En general, nos gustaría hallar un estadístico suficiente que reduzca los datos de la muestra tanto como sea posible. Aun cuando muchos estadísticos son suficientes para el parámetro θ asociado con una distribución específica, la aplicación del criterio de factorización lleva por lo general a un estadístico que proporciona el "mejor" resumen de la información en los datos. En el Ejemplo 9.5 este estadístico es $\overline{Y}$ (o alguna función biunívoca de él). En la siguiente sección, demostraremos la forma en que estos estadísticos suficientes se pueden usar para generar estimadores insesgados con varianza mínima.

Ejercicios

9.37 Suponga que $X_1, X_2, \ldots, X_n$ denotan n variables aleatorias de *Bernoulli* independientes e idénticamente distribuidas, tales que

$$P(X_i = 1) = p \quad \text{y} \quad P(X_i = 0) = 1 - p,$$

para cada $i = 1, 2, \ldots, n$. Demuestre que $\sum_{i=1}^{n} X_i$ es suficiente para p usando para ello el criterio de factorización dado en el Teorema 9.4.

9.38 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2.

a Si μ es desconocida y σ^2 es conocida, demuestre que $\overline{Y}$ es suficiente para μ.

b Si μ es conocida y σ^2 es desconocida, demuestre que $\sum_{i=1}^{n}(Y_i - \mu)^2$ es suficiente para σ^2.

c Si μ y σ^2 son desconocidas ambas, demuestre que $\sum_{i=1}^{n} Y_i$ y $\sum_{i=1}^{n} Y_i^2$ son conjuntamente suficientes para μ y σ^2. [Así, se deduce que $\overline{Y}$ y $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ o $\overline{Y}$ y S^2 son también conjuntamente suficientes para μ y σ^2.]

9.39 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Poisson con parámetro λ. Demuestre mediante condiciones que $\sum_{i=1}^{n} Y_i$ es suficiente para λ.

9.40 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Rayleigh con parámetro θ. (Consulte el Ejercicio 9.34.) Demuestre que $\sum_{i=1}^{n} Y_i^2$ es suficiente para θ.

9.41 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Weibull con m conocida y α desconocida. (Consulte el Ejercicio 6.26.) Demuestre que $\sum_{i=1}^{n} Y_i^m$ es suficiente para α.

9.42 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución geométrica con parámetro p, demuestre que $\overline{Y}$ es suficiente para p.

9.43 Si $Y_1, Y_2, \ldots, Y_n$ denotan variables aleatorias independientes y distribuidas de manera idéntica que pertenecen a una familia de distribución de potencias con parámetros α y θ; entonces, por el resultado del Ejercicio 6.17, si $\alpha, \theta > 0$,

$$f(y \mid \alpha, \theta) = \begin{cases} \alpha y^{\alpha-1}/\theta^{\alpha}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Si θ es conocida, demuestre que $\prod_{i=1}^{n} Y_i$ es suficiente para α.

9.44 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes y distribuidas idénticamente de una distribución de Pareto con parámetros α y β. Entonces, por el resultado del Ejercicio 6.18, si $\alpha, \beta > 0$,

$$f(y \mid \alpha, \beta) = \begin{cases} \alpha\beta^{\alpha} y^{-(\alpha+1)}, & y \ge \beta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Si β es conocida, demuestre que $\prod_{i=1}^{n} Y_i$ es suficiente para α.

9.45 Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una función de densidad de probabilidad de la familia exponencial (un parámetro) de modo que

$$f(y \mid \theta) = \begin{cases} a(\theta)b(y)e^{-[c(\theta)d(y)]}, & a \le y \le b, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde a y b no dependen de θ. Demuestre que $\sum_{i=1}^{n} d(Y_i)$ es suficiente para θ.

9.46 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución exponencial con media β, demuestre que $f(y \mid \beta)$ pertenece a la familia exponencial y que $\overline{Y}$ es suficiente para β.

9.47 Consulte el Ejercicio 9.43. Si θ es conocida, demuestre que la familia de distribuciones de potencias pertenece a la familia exponencial. ¿Cuál es un estadístico suficiente para α? ¿Contradice esto su respuesta al Ejercicio 9.43?

9.48 Consulte el Ejercicio 9.44. Si β es conocida, demuestre que la distribución de Pareto pertenece a la familia exponencial. ¿Cuál es un estadístico suficiente para α? Argumente que no hay contradicción entre su respuesta a este ejercicio y la respuesta que encontró en el Ejercicio 9.44.

***9.49** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución uniforme en el intervalo $(0, \theta)$. Demuestre que $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

***9.50** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la distribución uniforme en el intervalo (θ_1, θ_2). Demuestre que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ y $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ son conjuntamente suficientes para θ_1 y θ_2.

***9.51** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y \mid \theta) = \begin{cases} e^{-(y-\theta)}, & y \ge \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

***9.52** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

***9.53** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

***9.54** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan variables aleatorias independientes y distribuidas idénticamente de una familia de distribución de potencias con parámetros α y θ. Entonces, como en el Ejercicio 9.43, si $\alpha, \theta > 0$,

$$f(y \mid \alpha, \theta) = \begin{cases} \alpha y^{\alpha-1}/\theta^{\alpha}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $\text{máx}(Y_1, Y_2, \ldots, Y_n)$ y $\prod_{i=1}^{n} Y_i$ son conjuntamente suficientes para α y θ.

***9.55** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan variables aleatorias independientes y distribuidas idénticamente de una distribución de Pareto con parámetros α y β. Entonces, como en el Ejercicio 9.44, si $\alpha, \beta > 0$,

$$f(y \mid \alpha, \beta) = \begin{cases} \alpha\beta^{\alpha} y^{-(\alpha+1)}, & y \ge \beta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $\prod_{i=1}^{n} Y_i$ y $\text{mín}(Y_1, Y_2, \ldots, Y_n)$ son conjuntamente suficientes para α y β.

9.5 Teorema de Rao–Blackwell y estimación insesgada de varianza mínima

Los estadísticos suficientes desempeñan un importante papel para determinar buenos estimadores para parámetros. Si $\hat{\theta}$ es un estimador insesgado para θ y si U es un estadístico suficiente para θ, entonces hay una función de U que también es un estimador insesgado para θ y tiene una varianza *no mayor* que $\hat{\theta}$. Si buscamos estimadores insesgados con varianzas pequeñas, podemos restringir nuestra búsqueda a estimadores que sean funciones de estadísticos suficientes. La base teórica para las observaciones anteriores se proporciona en el siguiente resultado, conocido como el *teorema de Rao–Blackwell.*

TEOREMA **9.5**

El Teorema de Rao–Blackwell Sea $\hat{\theta}$ un estimador insesgado para θ tal que $V(\hat{\theta}) < \infty$. Si U es un estadístico suficiente para θ, definamos $\hat{\theta}^* = E(\hat{\theta} \mid U)$. Entonces, para toda θ,

$$E\left(\hat{\theta}^*\right) = \theta \quad \text{y} \quad V\left(\hat{\theta}^*\right) \le V(\hat{\theta}).$$

Demostración

Como U es suficiente para θ, la distribución condicional de cualquier estadístico (incluyendo $\hat{\theta}$), dada U, no depende de θ. Entonces, $\hat{\theta}^* = E(\hat{\theta} \mid U)$ no es una función de θ y es por tanto un estadístico.

Recuerde los Teoremas 5.14 y 5.15, donde consideramos la forma de hallar medias y varianzas de variables aleatorias con el uso de medias y varianzas condicionales. Como $\hat{\theta}$ es un estimador insesgado para θ, el Teorema 5.14 implica que

$$E(\hat{\theta}^*) = E[E(\hat{\theta} \mid U)] = E(\hat{\theta}) = \theta.$$

Entonces, $\hat{\theta}^*$ es un estimador insesgado para θ.

El Teorema 5.15 implica que

$$\begin{aligned} V(\hat{\theta}) &= V[E(\hat{\theta} \mid U)] + E[V(\hat{\theta} \mid U)] \\ &= V(\hat{\theta}^*) + E[V(\hat{\theta} \mid U)]. \end{aligned}$$

Como $V(\hat{\theta} \mid U = u) \geq 0$ para toda u, se deduce que $E[V(\hat{\theta} \mid U)] \geq 0$ y por tanto que $V(\hat{\theta}) \geq V(\hat{\theta}^*)$, como dijimos.

El Teorema 9.5 implica que un estimador insesgado para θ con una varianza pequeña es, o puede hacerse que sea, una función de un estadístico suficiente. Si tenemos un estimador insesgado para θ, podríamos mejorarlo con el uso del resultado del Teorema 9.5. Inicialmente puede parecer que si se aplica una vez el teorema de Rao–Blackwell para obtener un mejor estimador insesgado y luego se aplica nuevamente al nuevo estimador resultante se obtiene un estimador insesgado aún mejor. Si aplicamos el teorema de Rao–Blackwell usando el estadístico suficiente U, entonces $\hat{\theta}^* = E(\hat{\theta} \mid U)$ será una función del estadístico U, por ejemplo $\hat{\theta}^* = h(U)$. Suponga que reaplicamos el teorema de Rao–Blackwell a $\hat{\theta}^*$ con el uso del mismo estadístico suficiente U. Puesto que, en general, $E(h(U)|U) = h(U)$, vemos que usando de nuevo el teorema de Rao–Blackwell nuestro "nuevo" estimador es simplemente $h(U) = \hat{\theta}^*$. Esto es, si usamos el mismo estadístico suficiente en aplicaciones sucesivas del teorema de Rao–Blackwell, no ganamos nada después de la primera aplicación. La única forma en que aplicaciones sucesivas pueden llevar a mejores estimadores insesgados es si usamos un estadístico suficiente distinto cada vez que se reaplica el teorema. Así, no es necesario usar el teorema de Rao–Blackwell sucesivamente si usamos el estadístico suficiente correcto en nuestra aplicación inicial.

Debido a que numerosos estadísticos son suficientes para un parámetro θ asociado con una distribución, ¿qué estadístico suficiente debemos usar cuando apliquemos este teorema? Para las distribuciones que estudiamos en este texto, el criterio de factorización de manera típica identifica un estadístico U que mejor resume la información de los datos acerca del parámetro θ. Tales estadísticos reciben el nombre de *estadísticos suficientes mínimos*. El Ejercicio 9.66 presenta un método para determinar un estadístico suficiente mínimo que podría ser de interés para algunos lectores. En unos pocos de los ejercicios subsiguientes veremos que este método por lo general da los mismos estadísticos suficientes que los obtenidos con el criterio de factorización. En los casos que consideramos, estos estadísticos poseen otra propiedad (completabilidad) que garantiza que, si aplicamos el Teorema 9.5 usando U, no sólo obtenemos un estimador con una varianza más pequeña sino que también obtenemos en realidad un estimador insesgado para θ con *varianza mínima*. Este estimador recibe el nombre de *estimador insesgado de varianza mínima* (MVUE, por sus siglas en inglés). Puede consultar más detalles en la obra de Casella y Berger (2002), Hogg, Craig y McKean (2005), o Mood, Graybill y Boes (1974).

Por tanto, si empezamos con un estimador insesgado para un parámetro θ y el estadístico suficiente obtenido por medio del criterio de factorización, la aplicación del teorema de Rao–Blackwell en general lleva a un MVUE para el parámetro. El cálculo directo de los valores esperados condicionales puede ser difícil. No obstante, si U es el estadístico suficiente que mejor resume los datos y alguna función de U, por ejemplo $h(U)$, se puede hallar de modo que $E[h(U)] = \theta$, se deduce que $h(U)$ es el MVUE para θ. Ilustramos este método con varios ejemplos.

EJEMPLO **9.6** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución donde $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$, con p desconocida (es frecuente que tales variables aleatorias se denominen variables de *Bernoulli*). Use el criterio de factorización para hallar un estadístico suficiente que mejor resuma los datos. Proporcione un MVUE para p.

Solución Observe que la función de probabilidad anterior se puede escribir como

$$P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1.$$

Por tanto, la verosimilitud $L(p)$ es

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid p) &= p(y_1, y_2, \ldots, y_n \mid p) \\ &= p^{y_1}(1-p)^{1-y_1} \times p^{y_2}(1-p)^{1-y_2} \times \cdots \times p^{y_n}(1-p)^{1-y_n} \\ &= \underbrace{p^{\sum y_i}(1-p)^{n-\sum y_i}}_{g(\sum y_i,\, p)} \times \underbrace{1}_{h(y_1, y_2, \ldots, y_n)}. \end{aligned}$$

De acuerdo con el criterio de factorización, $U = \sum_{i=1}^{n} Y_i$ es suficiente para p. Este estadístico resume mejor la información acerca del parámetro p. Observe que $E(U) = np$, o bien, de la misma manera, $E(U/n) = p$. Así $U/n = \overline{Y}$ es un estimador insesgado para p. Como este estimador es una función del estadístico suficiente $\sum_{i=1}^{n} Y_i$, el estimador $\hat{p} = \overline{Y}$ es el MVUE para p. ■

EJEMPLO **9.7** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de la función de densidad de Weibull, dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre un MVUE para θ.

Solución Comencemos utilizando el criterio de factorización para hallar el estadístico suficiente que mejor resume la información acerca de θ.

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) \\ &= \left(\frac{2}{\theta}\right)^n (y_1 \times y_2 \times \cdots \times y_n) \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} y_i^2\right) \\ &= \underbrace{\left(\frac{2}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} y_i^2\right)}_{g(\sum y_i^2,\, \theta)} \times \underbrace{(y_1 \times y_2 \times \cdots \times y_n)}_{h(y_1, y_2, \ldots, y_n)}. \end{aligned}$$

Por tanto, $U = \sum_{i=1}^{n} Y_i^2$ es el estadístico suficiente mínimo para θ.

Ahora debemos determinar una función de este estadístico que sea insesgada para θ. Si hacemos $W = Y_i^2$, tenemos

$$f_W(w) = f(\sqrt{w})\frac{d(\sqrt{w})}{dw} = \left(\frac{2}{\theta}\right)\left(\sqrt{w}e^{-w/\theta}\right)\left(\frac{1}{2\sqrt{w}}\right) = \left(\frac{1}{\theta}\right)e^{-w/\theta}, \qquad w > 0.$$

Esto es, Y_i^2 tiene una distribución exponencial con parámetro θ. Como

$$E(Y_i^2) = E(W) = \theta \quad \text{y} \quad E\left(\sum_{i=1}^{n} Y_i^2\right) = n\theta,$$

se deduce que

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

es un estimador insesgado de θ que es una función del estadístico suficiente $\sum_{i=1}^{n} Y_i^2$. Por tanto, $\hat{\theta}$ es un MVUE del parámetro θ de Weibull. ■

El siguiente ejemplo ilustra la aplicación de esta técnica para estimar dos parámetros desconocidos.

EJEMPLO **9.8** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución normal con media desconocida μ y varianza σ^2. Encuentre los MVUE para μ y σ^2.

Solución De nuevo, al aplicar la función de verosimilitud, tenemos

$$\begin{aligned}
&L(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)\\
&\quad = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)\\
&\quad = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)\\
&\quad = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i + n\mu^2\right)\right]\\
&\quad = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\frac{-n\mu^2}{2\sigma^2}\right)\exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i\right)\right].
\end{aligned}$$

Por tanto, $\sum_{i=1}^{n} Y_i$ y $\sum_{i=1}^{n} Y_i^2$, constituyen conjuntamente estadísticos suficientes para μ y σ^2.

Sabemos, de cálculos anteriores, que $\overline{Y}$ es insesgado para μ y que

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]$$

es insesgado para σ^2. Debido a que estos estimadores son funciones de los estadísticos que mejor resumen la información acerca de μ y σ^2, son los MVUE para μ y σ^2. ■

El criterio de factorización también se puede utilizar, junto con el teorema de Rao–Blackwell, para hallar los MVUE para funciones de los parámetros asociados con una distribución. Ilustramos la técnica en el siguiente ejemplo.

EJEMPLO 9.9 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad exponencial dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta}\right) e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre un MVUE de $V(Y_i)$.

Solución En el Capítulo 4 determinamos que $E(Y_i) = \theta$ y que $V(Y_i) = \theta^2$. El criterio de factorización implica que $\sum_{i=1}^{n} Y_i$ es el mejor estadístico suficiente para θ. De hecho, $\overline{Y}$ es el MVUE de θ. Por tanto, podríamos usar $\overline{Y}^2$ como estimador de θ^2. Pero

$$E\left(\overline{Y}^2\right) = V(\overline{Y}) + [E(\overline{Y})]^2 = \frac{\theta^2}{n} + \theta^2 = \left(\frac{n+1}{n}\right)\theta^2.$$

Se deduce que $\overline{Y}^2$ es una estimación sesgada para θ^2. No obstante,

$$\left(\frac{n}{n+1}\right)\overline{Y}^2$$

es un MVUE de θ^2 porque es un estimador insesgado para θ^2 y una función del estadístico suficiente. Ningún otro estimador insesgado de θ^2 tendrá una varianza más pequeña que éste. ■

Con frecuencia un estadístico suficiente de un parámetro θ se puede usar para construir un intervalo de confianza exacto para θ si es que se puede hallar la distribución de probabilidad del estadístico. Generalmente los intervalos resultantes son los más cortos que se pueden hallar con un coeficiente de confianza especificado. Ilustramos la técnica con un ejemplo que comprende la distribución de Weibull.

EJEMPLO 9.10 El siguiente conjunto de datos, con mediciones en cientos de horas, representa la vida útil de diez componentes electrónicos idénticos que operan en un sistema de control de guía para proyectiles:

.637	1.531	.733	2.256	2.364
1.601	.152	1.826	1.868	1.126

Se supone que la vida útil de un componente de este tipo sigue una distribución de Weibull con una función de densidad dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Utilice los datos para construir un intervalo de confianza de 95% para θ.

Solución En el Ejemplo 9.7 vimos que el estadístico suficiente que resume mejor la información acerca de θ es $\sum_{i=1}^{n} Y_i^2$. Usaremos este estadístico para formar una cantidad pivote para construir el intervalo de confianza que buscamos.

Recuerde del Ejemplo 9.7 que $W_i = Y_i^2$ tiene una distribución exponencial con media θ. Ahora considere la transformación $T_i = 2W_i/\theta$. Entonces

$$f_T(t) = f_W\left(\frac{\theta t}{2}\right)\frac{d(\theta t/2)}{dt} = \left(\frac{1}{\theta}\right)e^{-(\theta t/2)/\theta}\left(\frac{\theta}{2}\right) = \left(\frac{1}{2}\right)e^{-t/2}, \qquad t > 0.$$

Así, para cada $i = 1, 2, \ldots, n$, T_i tiene una distribución χ^2 con 2 grados de libertad. Además, como las variables Y_i son independientes, las variables T_i son independientes para $i = 1, 2, \ldots, n$. La suma de variables aleatorias χ^2 independientes tiene una distribución χ^2 con grados de libertad iguales a la suma de los grados de libertad de las variables en la suma. Por tanto, la cantidad

$$\sum_{i=1}^{10} T_i = \frac{2}{\theta}\sum_{i=1}^{10} W_i = \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2$$

tiene una distribución χ^2 con 20 grados de libertad. Así que,

$$\frac{2}{\theta}\sum_{i=1}^{10} Y_i^2$$

es una cantidad pivote y podemos usar el método del pivote (Sección 8.5) para construir el intervalo de confianza buscado.

De la Tabla 6, Apéndice 3, podemos encontrar dos números a y b tales que

$$P\left(a \le \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2 \le b\right) = .95.$$

Si despejamos θ y la colocamos en la parte media, tenemos

$$\begin{aligned} .95 &= P\left(a \le \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2 \le b\right) = P\left(\frac{1}{b} \le \frac{\theta}{2\sum_{i=1}^{10} Y_i^2} \le \frac{1}{a}\right) \\ &= P\left(\frac{2\sum_{i=1}^{10} Y_i^2}{b} \le \theta \le \frac{2\sum_{i=1}^{10} Y_i^2}{a}\right). \end{aligned}$$

De la Tabla 6, Apéndice 3, el valor que limita un área de .025 en la cola inferior de la distribución χ^2 con 20 grados de libertad es $a = 9.591$. El valor que limita un área de .025 en la cola superior de la misma distribución es $b = 34.170$. Para los datos anteriores, $\sum_{i=1}^{10} Y_i^2 = 24.643$. Por tanto, el intervalo de confianza de 95% para el parámetro θ de Weibull es

$$\left(\frac{2(24.643)}{34.170}, \frac{2(24.643)}{9.591}\right), \quad \text{o} \quad (1.442, 5.139).$$

Éste es un intervalo bastante amplio para θ, pero está basado sólo en diez observaciones. ■

En esta sección hemos visto que el teorema de Rao–Blackwell implica que los estimadores insesgados con pequeñas varianzas son funciones de estadísticos suficientes. En general, el criterio de factorización presentado en la Sección 9.4 se puede aplicar para encontrar estadísticos suficientes que resuman mejor la información contenida en datos muestrales relacionada con los parámetros de interés. Para las distribuciones que consideramos en este texto, un MVUE para un parámetro θ objetivo se puede hallar de la siguiente manera. Primero, determinamos el mejor estadístico suficiente, U, en seguida hallamos una función de U, $h(U)$, tal que $E[h(U)] = \theta$.

Regularmente este método funciona bien, pero en ocasiones el mejor estadístico suficiente es una función muy complicada de las variables aleatorias observables en la muestra. En casos como éste puede ser difícil hallar una función del estadístico suficiente que sea un estimador insesgado para el parámetro objetivo. Por esta razón, dos métodos adicionales para determinar estimadores —el método de momentos y el método de máxima verosimilitud— se presentan en las dos secciones siguientes. Un tercer método de estimación importante, el método de mínimos cuadrados, es el tema del Capítulo 11.

Ejercicios

9.56 Consulte el Ejercicio 9.38(b). Encuentre un MVUE de σ^2.

9.57 Consulte el Ejercicio 9.18. ¿El estimador de σ^2 dado ahí es un MVUE de σ^2?

9.58 Consulte el Ejercicio 9.40. Use $\sum_{i=1}^{n} Y_i^2$ para hallar un MVUE de θ.

9.59 El número de descomposturas Y por día para cierta máquina es una variable aleatoria de Poisson con media λ. El costo diario de reparación de estas descomposturas está dado por $C = 3Y^2$. Si $Y_1, Y_2, \ldots, Y_n$ denota el número observado de descomposturas para n días seleccionados de manera independiente, encuentre un MVUE para $E(C)$.

9.60 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1,\ \theta > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Demuestre que esta función de densidad pertenece a la familia exponencial (de un parámetro) y que $\sum_{i=1}^{n} -\ln(Y_i)$ es suficiente para θ. (Véase Ejercicio 9.45.)

b Si $W_i = -\ln(Y_i)$, demuestre que W_i tiene una distribución exponencial con media $1/\theta$.

c Utilice métodos similares a los del Ejemplo 9.10 para demostrar que $2\theta \sum_{i=1}^{n} W_i$ tiene una distribución χ^2 con $2n$ grados de libertad.

d Demuestre que

$$E\left(\frac{1}{2\theta \sum_{i=1}^{n} W_i}\right) = \frac{1}{2(n-1)}.$$

[*Sugerencia:* recuerde el Ejercicio 4.112.]

e ¿Cuál es el MVUE para θ?

9.61 Consulte el Ejercicio 9.49. Use (Y_n) para hallar un MVUE de θ. (Véase el Ejemplo 9.1.)

9.62 Consulte el Ejercicio 9.51. Encuentre una función de $Y_{(1)}$ que sea un MVUE para θ.

9.63 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el Ejercicio 9.52 se demostró que $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

a Demuestre que $Y_{(n)}$ tiene función de densidad de probabilidad

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

b Encuentre el MVUE de θ.

9.64 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución normal con media μ y varianza 1.

a Demuestre que el MVUE de μ^2 es $\widehat{\mu^2} = \overline{Y}^2 - 1/n$.

b Obtenga la varianza de $\widehat{\mu^2}$.

***9.65** En este ejercicio ilustramos la aplicación directa del teorema de Rao–Blackwell. Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes de Bernoulli con

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1.$$

Esto es, $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$. Encuentre el MVUE de $p(1-p)$, que es un término de la varianza de Y_i o $W = \sum_{i=1}^{n} Y_i$, mediante los siguientes pasos.

a Sea

$$T = \begin{cases} 1, & \text{si } Y_1 = 1 \text{ y } Y_2 = 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $E(T) = p(1-p)$.

b Demuestre que

$$P(T = 1 \mid W = w) = \frac{w(n-w)}{n(n-1)}.$$

c Demuestre que

$$E(T \mid W) = \frac{n}{n-1}\left[\frac{W}{n}\left(1 - \frac{W}{n}\right)\right] = \frac{n}{n-1}\overline{Y}(1-\overline{Y})$$

y por tanto que $n\overline{Y}(1-\overline{Y})/(n-1)$ es el MVUE de $p(1-p)$.

***9.66** La función de verosimilitud $L(y_1, y_2, \ldots, y_n \mid \theta)$ toma valores diferentes dependiendo de los argumentos $(y_1, y_2, \ldots, y_n)$. Un método para deducir un estadístico suficiente *mínimo* desarrollado por Lehmann y Scheffé usa la relación entre las verosimilitudes evaluadas en dos puntos $(x_1, x_2, \ldots, x_n)$ y $(y_1, y_2, \ldots, y_n)$:

$$\frac{L(x_1, x_2, \ldots, x_n \mid \theta)}{L(y_1, y_2, \ldots, y_n \mid \theta)}.$$

Muchas veces es posible hallar una función $g(x_1, x_2, \ldots, x_n)$ tal que esta relación no tenga parámetros θ desconocidos si y sólo si $g(x_1, x_2, \ldots, x_n) = g(y_1, y_2, \ldots, y_n)$. Si se puede hallar esa función g, entonces $g(Y_1, Y_2, \ldots, Y_n)$ es un estadístico suficiente mínimo para θ.

a Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Bernoulli (vea el Ejemplo 9.6 y el Ejercicio 9.65) con p desconocida.

i Demuestre que

$$\frac{L(x_1, x_2, \ldots, x_n \mid p)}{L(y_1, y_2, \ldots, y_n \mid p)} = \left(\frac{p}{1-p}\right)^{\Sigma x_i - \Sigma y_i}.$$

ii Demuestre que para que esta relación sea independiente de p, debemos tener

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \quad \text{o} \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i .$$

iii De acuerdo con el método de Lehmann y Scheffé, ¿cuál es el estadístico suficiente mínimo para p? ¿Cómo se compara este estadístico suficiente con el estadístico suficiente deducido en el Ejemplo 9.6 usando el criterio de factorización?

b Considere la densidad de Weibull estudiada en el Ejemplo 9.7.

i Demuestre que

$$\frac{L(x_1, x_2, \ldots, x_n \mid \theta)}{L(y_1, y_2, \ldots, y_n \mid \theta)} = \left(\frac{x_1 x_2 \cdots x_n}{y_1 y_2 \cdots y_n}\right) \exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right)\right].$$

ii Demuestre que $\sum_{i=1}^{n} Y_i^2$ es un estadístico suficiente mínimo para θ.

***9.67** Consulte el Ejercicio 9.66. Suponga que se toma una muestra de tamaño n de una población normal con media μ y varianza σ^2. Demuestre que $\sum_{i=1}^{n} Y_i$, y $\sum_{i=1}^{n} Y_i^2$ conjuntamente forman estadísticos suficientes mínimos para μ y σ^2.

***9.68** Suponga que un estadístico U tiene una función de densidad de probabilidad que es positiva en el intervalo $a \le u \le b$ y suponga que la densidad depende de un parámetro θ que puede variar en el intervalo $\alpha_1 \le \theta \le \alpha_2$. Suponga también que $g(u)$ es continua para u en el intervalo $[a, b]$. Si $E[(g(U) \mid \theta] = 0$ para toda θ en el intervalo $[\alpha_1, \alpha_2]$ implica que $g(u)$ sea idénticamente cero, entonces se dice que la familia de funciones de densidad $\{f_U(u \mid \theta), \alpha_1 \le \theta \le \alpha_2\}$ está *completa*. (Todos los estadísticos que empleamos en la Sección 9.5 tienen familias completas de funciones de densidad.) Suponga que U es un estadístico suficiente para θ, y $g_1(U)$ y $g_2(U)$ son estimadores insesgados de θ. Demuestre que, si la familia de funciones de densidad para U está completa, $g_1(U)$ debe ser igual a $g_2(U)$, y entonces hay una función *única* de U que es un estimador insesgado de θ.

Acoplada con el teorema de Rao–Blackwell, la propiedad de completabilidad de $f_U(u \mid \theta)$, junto con la suficiencia de U, asegura que hay un estimador insesgado único de varianza mínima (UMVUE) de θ.

9.6 Método de momentos

En esta sección estudiaremos uno de los métodos más antiguos para obtener estimadores puntuales: el método de momentos. Un método más refinado, el de verosimilitud máxima, es el tema de la Sección 9.7.

El método de momentos es un procedimiento muy sencillo para hallar un estimador para uno o más parámetros poblacionales. Recuerde que el k-ésimo momento de una variable aleatoria, tomado alrededor del origen, es

$$\mu_k' = E(Y^k).$$

El correspondiente k-ésimo momento muestral es el promedio

$$m_k' = \frac{1}{n}\sum_{i=1}^{n} Y_i^k .$$

El método de momentos está basado en la idea de que los momentos muestrales deben dar buenas estimaciones de los momentos poblacionales correspondientes.

Es decir, m'_k debe ser un buen estimador de μ'_k, para $k = 1, 2, \ldots$ Entonces, debido a que los momentos poblacionales $\mu'_1, \mu'_2, \ldots, \mu'_k$ son funciones de los parámetros poblacionales, podemos igualar los correspondientes momentos poblacionales y muestrales y despejar los estimadores deseados. En consecuencia, el método de momentos se puede expresar como sigue.

Método de momentos

Escoja como estimaciones los valores de los parámetros que son soluciones de las ecuaciones $\mu'_k = m'_k$, para $k = 1, 2, \ldots, t$, donde t es el número de parámetros por estimar.

EJEMPLO **9.11** Una muestra aleatoria de n observaciones, $Y_1, Y_2, \ldots, Y_n$ se selecciona de una población en la que Y_i, para $i = 1, 2, \ldots, n$, posee una función de densidad de probabilidad uniforme en el intervalo $(0, \theta)$ donde θ es desconocida. Use el método de momentos para estimar el parámetro θ.

Solución El valor de μ'_1 para una variable aleatoria uniforme es

$$\mu'_1 = \mu = \frac{\theta}{2}.$$

El correspondiente primer momento muestral es

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

Igualando la población correspondiente y el momento muestral, obtenemos

$$\mu'_1 = \frac{\theta}{2} = \overline{Y}.$$

El estimador que se obtiene mediante el método de momentos para θ es la solución de la ecuación anterior. Esto es, $\hat{\theta} = 2\overline{Y}$. ■

Para las distribuciones que consideramos en este texto, los métodos de la Sección 9.3 se pueden utilizar para demostrar que los momentos muestrales son estimadores consistentes de los momentos poblacionales correspondientes. Debido a que los estimadores obtenidos con el método de momentos obviamente son funciones de los momentos muestrales, suelen ser estimadores consistentes de sus respectivos parámetros.

EJEMPLO **9.12** Demuestre que el estimador $\hat{\theta} = 2\overline{Y}$, encontrado en el Ejemplo 9.11, es un estimador consistente para θ.

Solución En el Ejemplo 9.1, demostramos que $\hat{\theta} = 2\overline{Y}$ es un estimador insesgado para θ y que $V(\hat{\theta}) = \theta^2/3n$. Como $\lim_{n\to\infty} V(\hat{\theta}) = 0$, el Teorema 9.1 implica que $\hat{\theta} = 2\overline{Y}$ es un estimador consistente para θ. ■

Aun cuando el estimador $\hat{\theta}$ obtenido en el Ejemplo 9.11 es consistente, no es necesariamente el mejor estimador para θ. De hecho, el criterio de factorización da $Y_{(n)} = \text{máx}(Y_1, Y_2, \dots, Y_n)$ como el mejor estadístico suficiente para θ. Entonces, de acuerdo con el teorema de Rao–Blackwell, el estimador que se obtenga mediante el método de momentos tendrá varianza más grande que un estimador insesgado basado en $Y_{(n)}$. De hecho, se demostró que este es el caso en el Ejemplo 9.1.

EJEMPLO **9.13** Una muestra aleatoria de n observaciones $Y_1, Y_2, \dots, Y_n$, se selecciona de una población en la que Y_i, para $i = 1, 2, \dots, n$, posee una función de densidad de probabilidad gamma con parámetros α y β (véase la Sección 4.6 para la función de densidad de probabilidad gamma). Encuentre los estimadores por el método de momentos para los parámetros desconocidos α y β.

Solución Debido a que buscamos estimadores para dos parámetros α y β, debemos igualar dos pares de momentos poblacionales y muestrales.

Los primeros dos momentos de la distribución gamma con parámetros α y β son (si es necesario, vea al final de este libro)

$$\mu_1' = \mu = \alpha\beta \quad \text{y} \quad \mu_2' = \sigma^2 + \mu^2 = \alpha\beta^2 + \alpha^2\beta^2.$$

Ahora iguale estas cantidades con sus correspondientes momentos muestrales y despeje $\hat{\alpha}$ y $\hat{\beta}$. Así,

$$\mu_1' = \alpha\beta = m_1' = \overline{Y},$$

$$\mu_2' = \alpha\beta^2 + \alpha^2\beta^2 = m_2' = \frac{1}{n}\sum_{i=1}^{n} Y_i^2.$$

De la primera ecuación, obtenemos $\hat{\beta} = \overline{Y}/\hat{\alpha}$. Sustituyendo en la segunda ecuación y despejando $\hat{\alpha}$, obtenemos

$$\hat{\alpha} = \frac{\overline{Y}^2}{\left(\sum Y_i^2/n\right) - \overline{Y}^2} = \frac{n\overline{Y}^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}.$$

Sustituyendo $\hat{\alpha}$ en la primera ecuación, obtenemos

$$\hat{\beta} = \frac{\overline{Y}}{\hat{\alpha}} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n\overline{Y}}.$$

■

Los estimadores del método de momentos $\hat{\alpha}$ y $\hat{\beta}$ del Ejemplo 9.13 son consistentes. $\overline{Y}$ converge en probabilidad en $E(Y_i) = \alpha\beta$ y $(1/n)\sum_{i=1}^{n} Y_i^2$ converge en probabilidad en $E(Y_i^2) = \alpha\beta^2 + \alpha^2\beta^2$. Por lo tanto,

$$\hat{\alpha} = \frac{\overline{Y}^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2} \quad \text{es un estimador consistente de} \quad \frac{(\alpha\beta)^2}{\alpha\beta^2 + \alpha^2\beta^2 - (\alpha\beta)^2} = \alpha,$$

y

$$\hat{\beta} = \frac{\overline{Y}}{\hat{\alpha}} \quad \text{es un estimador consistente de} \quad \frac{\alpha\beta}{\alpha} = \beta.$$

Usando el criterio de factorización podemos demostrar que $\sum_{i=1}^{n} Y_i$ y el producto $\prod_{i=1}^{n} Y_i$ son estadísticos suficientes para la función de densidad gamma. Como los estimadores del método de momentos $\hat{\alpha}$ y $\hat{\beta}$ no son funciones de estos estadísticos suficientes, podemos hallar más estimadores eficientes para los parámetros α y β. No obstante, es considerablemente más difícil aplicar otros métodos para hallar estimadores de estos parámetros.

Para resumir, el método de momentos permite generar estimadores de parámetros desconocidos al igualar los correspondientes momentos muestrales y poblacionales. El método es fácil de emplear y proporciona estimadores consistentes, pero los estimadores obtenidos por este método en ocasiones no son funciones de estadísticos suficientes. En consecuencia, es frecuente que los estimadores del método de momentos no sean eficientes y en muchos casos sean sesgados. Las virtudes básicas de este método son su facilidad de aplicación y que a veces proporciona estimadores con propiedades razonables.

Ejercicios

9.69 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y \mid \theta) = \begin{cases} (\theta + 1)y^{\theta}, & 0 < y < 1 \quad \theta > -1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre un estimador para θ por el método de momentos. Demuestre que el estimador es consistente. ¿El estimador es una función del estadístico suficiente $-\sum_{i=1}^{n} \ln(Y_i)$ que podemos obtener del criterio de factorización? ¿Qué implicaciones tiene esto?

9.70 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución de Poisson con media λ. Encuentre el estimador del método de momentos para λ.

9.71 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de la distribución normal con media conocida $\mu = 0$ y varianza desconocida σ^2, encuentre el estimador de σ^2 por el método de momentos.

9.72 Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución normal con media μ y varianza σ^2, encuentre los estimadores de μ y σ^2 por medio del método de momentos.

9.73 Una urna contiene θ bolas negras y $N - \theta$ bolas blancas. Una muestra de n bolas se ha de seleccionar sin restitución. Sea Y el número de bolas negras de la muestra. Demuestre que $(N/n)Y$ es el estimador del método de momentos para θ.

9.74 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2}{\theta^2}\right)(\theta - y), & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre un estimador para θ usando el método de momentos.

b ¿Este estimador es un estadístico suficiente para θ?

9.75 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad dada por

$$f(y \mid \theta) = \begin{cases} \dfrac{\Gamma(2\theta)}{[\Gamma(\theta)]^2}(y^{\theta-1})(1 - y)^{\theta-1}, & 0 \le y \le 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre el estimador de θ por el método de momentos.

9.76 Sean $X_1, X_2, \ldots, X_n$ variables aleatorias de Bernoulli independientes tales que $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$ para cada $i = 1, 2, 3, \ldots$ Con la variable aleatoria Y denote el número de intentos necesario para obtener el primer éxito, es decir, el valor de i para el cual $X_i = 1$ ocurre primero. Entonces Y tiene una distribución geométrica con $P(Y = y) = (1 - p)^{y-1}p$, para $y = 1, 2, 3, \ldots$ Encuentre el estimador del método de momentos para p basado en esta única observación de Y.

9.77 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias uniformes independientes y distribuidas idénticamente en el intervalo $(0, 3\theta)$. Deduzca el estimador del método de momentos para θ.

9.78 Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes y distribuidas idénticamente de una familia de distribución de potencias con parámetros α y $\theta = 3$. Entonces, como en el Ejercicio 9.43, si $\alpha > 0$,

$$f(y|\alpha) = \begin{cases} \alpha y^{\alpha-1}/3^{\alpha}, & 0 \leq y \leq 3, \\ 0, & \text{en cualquier otro punto..} \end{cases}$$

Demuestre que $E(Y_1) = 3\alpha/(\alpha + 1)$ y deduzca el estimador del método de momentos para α.

***9.79** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes y distribuidas idénticamente de una distribución de Pareto con parámetros α y β, donde β es conocida. Entonces, si $\alpha > 0$,

$$f(y|\alpha, \beta) = \begin{cases} \alpha\beta^{\alpha} y^{-(\alpha+1)}, & y \geq \beta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que $E(Y_i) = \alpha\beta/(\alpha - 1)$ si $\alpha > 1$ y $E(Y_i)$ no está definida si $0 < \alpha < 1$. Entonces, el estimador del método de momentos para α no está definido.

9.7 Método de máxima verosimilitud

En la Sección 9.5 presentamos un método para obtener un estimador insesgado de varianza mínima (MVUE) por un parámetro objetivo: usando el criterio de factorización junto con el teorema de Rao–Blackwell. El método requiere que encontremos alguna función de un estadístico suficiente mínimo que es un estimador insesgado para el parámetro objetivo. Aun cuando tenemos un método para hallar un estadístico suficiente, la determinación de la función del estadístico suficiente mínimo que proporciona un estimador insesgado puede ser en gran medida una cuestión de azar. La Sección 9.6 contiene una exposición del método de momentos. El método de momentos es intuitivo y fácil de aplicar, pero por lo general no lleva a los mejores estimadores. En esta sección presentamos el método de máxima verosimilitud que con frecuencia proporciona estimadores insesgados de varianza mínima (MVUE).

Usamos un ejemplo para ilustrar la lógica en la que está basado el método de máxima verosimilitud. Suponga que tenemos una caja que contiene tres pelotas. Sabemos que cada una de las pelotas puede ser roja o blanca, pero no sabemos el número total de cualquiera de los colores. No obstante, podemos muestrear aleatoriamente dos de las pelotas sin restitución. Si nuestra muestra aleatoria contiene dos pelotas rojas, ¿cuál sería una buena estimación del número total de pelotas rojas en la caja? Obviamente, el número de pelotas rojas en la caja debe ser dos o tres (si hubiera cero o una pelota roja en la caja, sería imposible obtener dos pelotas rojas cuando se hace muestreo sin restitución). Si hay dos pelotas rojas y una pelota blanca en la caja, la probabilidad de seleccionar aleatoriamente dos pelotas rojas es

$$\frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = \frac{1}{3}.$$

Por otra parte, si hay tres pelotas rojas en la caja, la probabilidad de seleccionar aleatoriamente dos pelotas rojas es

$$\frac{\binom{3}{2}}{\binom{3}{2}} = 1.$$

Parece razonable escoger el tres como la estimación del número de pelotas rojas en la caja porque esta estimación *maximiza la probabilidad* de obtener la muestra observada. Desde luego que es posible que la caja contenga sólo dos pelotas rojas, pero el resultado observado confiere más crédito a que haya tres pelotas rojas en la caja.

Este ejemplo ilustra un método para hallar un estimador que puede aplicarse a cualquier situación. La técnica, llamada *método de máxima verosimilitud*, selecciona como estimaciones los valores de los parámetros que maximizan la verosimilitud (la función de probabilidad conjunta o función de densidad conjunta) de la muestra observada (vea la Definición 9.4). Recuerde que nos referimos a este método de estimación en el Capítulo 3 donde, en los Ejemplos 3.10 y 3.13 y en el Ejercicio 3.101, encontramos las estimaciones de máxima verosimilitud del parámetro p con base en observaciones individuales en variables aleatorias binomiales negativas, binomiales y geométricas, respectivamente.

Método de máxima verosimilitud

Suponga que la función de verosimilitud depende de k parámetros $\theta_1, \theta_2, \ldots, \theta_k$. Escoja como estimaciones los valores de los parámetros que maximicen la verosimilitud $L(y_1, y_2, \ldots, y_n \mid \theta_1, \theta_2, \ldots, \theta_k)$.

Para destacar el hecho de que la función de verosimilitud es una función de los parámetros $\theta_1, \theta_2, \ldots, \theta_k$, a veces expresamos la función de verosimilitud como $L(\theta_1, \theta_2, \ldots, \theta_k)$. Es común referirnos a estimadores de máxima verosimilitud como a los MLE, por sus siglas en inglés. Ilustramos el método con un ejemplo.

EJEMPLO **9.14** Un experimento binomial consistente en n ensayos resultó en las observaciones $y_1, y_2, \ldots, y_n$, donde $y_i = 1$ si el i-ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de p, la probabilidad de un éxito.

Solución La verosimilitud de la muestra observada es la probabilidad de observar $y_1, y_2, \ldots, y_n$. En consecuencia,

$$L(p) = L(y_1, y_2, \ldots, y_n \mid p) = p^y(1-p)^{n-y}, \qquad \text{donde } y = \sum_{i=1}^{n} y_i.$$

Ahora deseamos determinar el valor de p que maximice $L(p)$. Si $y = 0$, $L(p) = (1-p)^n$ y $L(p)$ se maximiza cuando $p = 0$. Análogamente, si $y = n$, $L(p) = p^n$ y $L(p)$ se maximiza cuando $p = 1$. Si $y = 1, 2, \ldots, n-1$, entonces $L(p) = p^y(1-p)^{1-y}$ es cero cuando $p = 0$ y $p = 1$ y es continua para valores de p entre 0 y 1. Entonces, para $y = 1, 2, \ldots, n-1$, podemos determinar el valor de p que maximice $L(p)$ al igualar a cero la derivada $dL(p)/dp$ y despejando p.

Usted notará que $\ln[L(p)]$ es una función creciente en forma monotónica de $L(p)$. En consecuencia, tanto $\ln[L(p)]$ como $L(p)$ se maximizan para el mismo valor de p. Como $L(p)$ es un producto de funciones de p y hallar la derivada de productos resulta laborioso, es más fácil hallar el valor de p que maximice $\ln[L(p)]$. Tenemos

$$\ln[L(p)] = \ln\left[p^y(1-p)^{n-y}\right] = y\ln p + (n-y)\ln(1-p).$$

Si $y = 1, 2, \ldots, n-1$, la derivada de $\ln[L(p)]$ con respecto a p, es

$$\frac{d\ \ln[L(p)]}{dp} = y\left(\frac{1}{p}\right) + (n-y)\left(\frac{-1}{1-p}\right).$$

Para $y = 1, 2, \ldots, n-1$, el valor de p que maximice (o minimice) $\ln[L(p)]$ es la solución de la ecuación

$$\frac{y}{\hat{p}} - \frac{n-y}{1-\hat{p}} = 0.$$

Resolviendo, obtenemos la estimación $\hat{p} = y/n$. Se puede verificar fácilmente que esta solución se presenta cuando $\ln[L(p)]$ [y por tanto $L(p)$] alcanza un máximo.

Debido a que $L(p)$ se maximiza en $p = 0$ cuando $y = 0$, en $p = 1$ cuando $y = n$ y en $p = y/n$ cuando $y = 1, 2, \ldots, n-1$, cualquiera que sea el valor observado de y, $L(p)$ se maximiza cuando $p = y/n$.

El MLE, $\hat{p} = Y/n$, es la fracción de éxitos en el número total de intentos n. Por tanto, el MLE de p es en realidad el estimador intuitivo para p que usamos en todo el Capítulo 8. ■

EJEMPLO 9.15 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. Encuentre los estimadores de máxima verosimilitud (MLE) de μ y σ^2.

Solución Como $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias continuas, $L(\mu, \sigma^2)$ es la densidad conjunta de la muestra. Así, $L(\mu, \sigma^2) = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)$. En este caso,

$$\begin{aligned}
L(\mu, \sigma^2) &= f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2) \\
&= f(y_1 \mid \mu, \sigma^2) \times f(y_2 \mid \mu, \sigma^2) \times \cdots \times f(y_n \mid \mu, \sigma^2) \\
&= \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_1-\mu)^2}{2\sigma^2}\right]\right\} \times \cdots \times \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_n-\mu)^2}{2\sigma^2}\right]\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right].
\end{aligned}$$

[Recuerde que $\exp(w)$ es sólo otra forma de escribir e^w.] Además,

$$\ln\left[L(\mu, \sigma^2)\right] = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2.$$

Los MLE de μ y σ^2 son los valores que hacen $\ln[L(\mu, \sigma^2)$ un máximo. Evaluando derivadas con respecto a μ y σ^2, obtenemos

$$\frac{\partial\{\ln[L(\mu, \sigma^2)]\}}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)$$

y

$$\frac{\partial\{\ln[L(\mu, \sigma^2)]\}}{\partial\sigma^2} = -\left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Igualando a cero estas derivadas y resolviendo simultáneamente, obtenemos de la primera ecuación

$$\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\mu}) = 0, \quad \text{o} \quad \sum_{i=1}^{n} y_i - n\hat{\mu} = 0, \quad \text{y} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}.$$

Sustituyendo $\overline{y}$ por $\hat{\mu}$ en la segunda ecuación y despejando $\hat{\sigma}^2$, tenemos

$$-\left(\frac{n}{\hat{\sigma}^2}\right) + \frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}(y_i - \overline{y})^2 = 0, \quad \text{o} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

Entonces, $\overline{Y}$ y $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ son los MLE de μ y σ^2, respectivamente. Observe que $\overline{Y}$ es insesgada para μ. Aun cuando $\hat{\sigma}^2$ no está insesgada para σ^2, se puede ajustar fácilmente al estimador insesgado S^2 (vea el Ejemplo 8.1). ■

EJEMPLO **9.16** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de observaciones de una distribución uniforme con función de densidad de probabilidad $f(y_i|\theta) = 1/\theta$, para $0 \le y_i \le \theta$ e $i = 1, 2, \ldots, n$. Encuentre el MLE de θ.

Solución En este caso, la verosimilitud está dada por

$$\begin{aligned} L(\theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) = f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta) \\ &= \begin{cases} \dfrac{1}{\theta} \times \dfrac{1}{\theta} \times \cdots \times \dfrac{1}{\theta} = \dfrac{1}{\theta^n}, & \text{si } 0 \le y_i \le \theta, i = 1, 2, \ldots, n, \\ 0, & \text{en cualquier otro punto.} \end{cases} \end{aligned}$$

Obviamente, $L(\theta)$ no está maximizado cuando $L(\theta) = 0$. Usted notará que $1/\theta^n$ es una función de θ que decrece en forma monotónica. Por tanto, en ninguna parte del intervalo $0 < \theta < \infty$ es $d[1/\theta^n]/d\theta$ igual a cero. No obstante, $1/\theta^n$ aumenta cuando θ disminuye y $1/\theta^n$ se maximiza al seleccionar θ tan pequeña como sea posible, sujeto a la restricción de que todos los valores de y_i estén entre 0 y θ. El valor más pequeño de θ que satisface esta restricción es la máxima observación del conjunto $y_1, y_2, \ldots, y_n$. Esto es, $\hat{\theta} = Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ es el MLE para θ. Este MLE para θ no es un estimador insesgado de θ, pero se puede ajustar para ser insesgado, como se muestra en el Ejemplo 9.1. ■

Hemos visto que los estadísticos suficientes que mejor resumen los datos tienen propiedades deseables y con frecuencia se pueden usar para determinar un estimador insesgado de varianza mínima (MVUE) para parámetros de interés. Si U es *cualquier* estadístico suficiente para la estimación de un parámetro θ, incluyendo el estadístico suficiente obtenido del uso óptimo del criterio de factorización, el MLE es siempre alguna función de U. Esto es, el MLE depende de las observaciones muestrales sólo mediante el valor de un estadístico suficiente.

Para demostrar esto, sólo necesitamos observar que si U es un estadístico suficiente para θ, el criterio de factorización (Teorema 9.4) implica que la verosimilitud puede ser factorizada como

$$L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta)\, h(y_1, y_2, \ldots, y_n),$$

donde $g(u, \theta)$ es una función de sólo u y θ y $h(y_1, y_2, \ldots, y_n)$ *no depende de* θ. Por tanto, se deduce que

$$\ln[L(\theta)] = \ln[g(u, \theta)] + \ln[h(y_1, y_2, \ldots, y_n)].$$

Observe que $\ln[h(y_1, y_2, \ldots, y_n)]$ no depende de θ y por tanto maximizar $\ln[L(\theta)]$ con respecto a θ es equivalente a maximizar $\ln[g(u, \theta)]$ con respecto a θ. Como $\ln[g(u, \theta)]$ depende de los datos sólo mediante el valor del estadístico suficiente U, el MLE para θ es siempre alguna función de U. En consecuencia, si un MLE para un parámetro se puede hallar y luego ajustar para ser insesgado, el estimador resultante es con frecuencia un MVUE del parámetro en cuestión.

Los MLE tienen algunas propiedades adicionales que hacen que este método de estimación sea particularmente atractivo. En el Ejemplo 9.9 consideramos la estimación de θ^2, una función del parámetro θ. Funciones de otros parámetros también pueden ser de interés. Por ejemplo, la varianza de una variable aleatoria binomial es $np(1-p)$, una función del parámetro p. Si Y tiene una distribución de Poisson con media λ, se deduce que $P(Y = 0) = e^{-\lambda}$; podemos desear calcular esta función de λ. En general, si θ es el parámetro asociado con una distribución, en ocasiones estamos interesados en calcular alguna función de θ, por ejemplo $t(\theta)$, en lugar de θ misma. En el Ejercicio 9.94 usted demostrará que si $t(\theta)$ es una función biunívoca de θ y si $\hat{\theta}$ es el MLE para θ, entonces el MLE de $t(\theta)$ está dado por

$$\widehat{t(\theta)} = t(\hat{\theta}).$$

Este resultado, a veces conocido como la *propiedad de invarianza* de los MLE, también se cumple para cualquier función de un parámetro de interés (no sólo funciones biunívocas). Véanse más detalles en la obra de Casella y Berger (2002).

EJEMPLO **9.17** En el Ejemplo 9.14, encontramos que el estimador de máxima verosimilitud (MLE) de una proporción binomial p está dado por $\hat{p} = Y/n$. ¿Cuál es el MLE para la varianza de Y?

Solución La varianza de una variable aleatoria binomial Y está dada por $V(Y) = np(1-p)$. Como $V(Y)$ es una función del parámetro binomial p, por ejemplo, $V(Y) = t(p)$ con $t(p) = np(1-p)$, se deduce que el MLE de $V(Y)$ está dado por

$$\widehat{V(Y)} = \widehat{t(p)} = t(\hat{p}) = n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right).$$

Este estimador no está insesgado, pero, usando el resultado del Ejercicio 9.65, podemos fácilmente ajustarlo para hacerlo insesgado. En realidad,

$$n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)\left(\frac{n}{n-1}\right) = \left(\frac{n^2}{n-1}\right)\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$$

es el estimador insesgado único de varianza mínima (UMVUE) para $t(p) = np(1-p)$. ■

En la siguiente sección (opcional), resumimos algunas de las propiedades útiles y convenientes de los MLE con muestras grandes.

Ejercicios

9.80 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de la distribución de Poisson con media λ.

a Encuentre el MLE $\hat{\lambda}$ para λ.

b Encuentre el valor esperado y la varianza de $\hat{\lambda}$.

c Demuestre que el estimador del inciso a es consistente para λ.

d ¿Cuál es el MLE para $P(Y = 0) = e^{-\lambda}$?

9.81 Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una población distribuida exponencialmente con media θ. Encuentre el MLE de la varianza poblacional θ^2. [*Sugerencia:* recuerde el Ejemplo 9.9.]

9.82 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta}\right) r y^{r-1} e^{-y^r/\theta}, & \theta > 0,\ y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde r es una constante positiva conocida.

a Encuentre un estadístico suficiente para θ.

b Encuentre el MLE de θ.

c ¿El estimador del inciso b es un MVUE para θ?

9.83 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución uniforme con función de densidad de probabilidad

$$f(y \mid \theta) = \begin{cases} \dfrac{1}{2\theta + 1}, & 0 \le y \le 2\theta + 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Obtenga el MLE de θ.

b Obtenga el MLE para la *varianza* de la distribución subyacente.

9.84 Cierto tipo de componente electrónico tiene una duración Y (en horas) con función de densidad de probabilidad dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta^2}\right) y e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Esto es, Y tiene una distribución gamma con parámetros $\alpha = 2$ y θ. Con $\hat{\theta}$ denote el MLE de θ. Suponga que tres de tales componentes, probados independientemente, tuvieron duraciones de 120, 130 y 128 horas.

a Encuentre el MLE de θ.

b Encuentre $E(\hat{\theta})$ y $V(\hat{\theta})$.

c Suponga que θ en realidad es igual a 130. Proporcione un límite aproximado que pudiera esperarse para el error de estimación.

d ¿Cuál es el MLE para la varianza de Y?

9.85 Sean $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad dada por

$$f(y \mid \alpha, \theta) = \begin{cases} \left(\dfrac{1}{\Gamma(\alpha)\theta^{\alpha}}\right) y^{\alpha-1} e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\alpha > 0$ es conocida.

a Encuentre el MLE $\hat{\theta}$ de θ.

b Encuentre el valor esperado y la varianza de $\hat{\theta}$.

c Demuestre que $\hat{\theta}$ es consistente para θ.

d ¿Cuál es el mejor (mínimo) estadístico suficiente para θ en este problema?

e Suponga que $n = 5$ y $\alpha = 2$. Use el mínimo estadístico suficiente para construir un intervalo de confianza de 90% para θ. [*Sugerencia:* transforme en una distribución χ^2.]

9.86 Suponga que $X_1, X_2, \ldots, X_m$, que representan la producción por acre para la variedad A de maíz, constituyen una muestra aleatoria de una distribución normal con media μ_1 y varianza σ^2. También, $Y_1, Y_2, \ldots, Y_n$, que representan la producción para la variedad B de maíz, constituyen una muestra aleatoria de una distribución normal con media μ_2 y varianza σ^2. Si las X y las Y son independientes, encuentre el MLE para la varianza común σ^2. Suponga que μ_1 y μ_2 son desconocidas.

9.87 Una muestra aleatoria de 100 votantes seleccionados de una población grande reveló que 30 están a favor del candidato A, 38 a favor del candidato B y 32 a favor del candidato C. Encuentre los MLE para las proporciones de votantes en la población que están a favor de los candidatos A, B y C, respectivamente. Calcule la diferencia entre las fracciones que están a favor de A y B y ponga un límite de desviación estándar de 2 en el error de estimación.

9.88 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad

$$f(y \mid \theta) = \begin{cases} (\theta + 1)y^{\theta}, & 0 < y < 1,\ \theta > -1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre el MLE para θ. Compare su respuesta con el estimador del método de momentos hallado en el Ejercicio 9.69.

9.89 Se sabe que la probabilidad p de obtener una cara al lanzar al aire una moneda desbalanceada es $1/4$ o $3/4$. La moneda es lanzada dos veces al aire y se observa un valor para Y, el número de caras. Para cada valor posible de Y, ¿cuál de los dos valores para p ($1/4$ o $3/4$) maximiza la probabilidad de que $Y = y$? Dependiendo del valor de y observado realmente, ¿cuál es el MLE de p?

9.90 Veinticinco hombres que forman parte de una muestra aleatoria de 100 hombres están a favor de una controvertida propuesta. De una muestra aleatoria independiente de 100 mujeres, un total de 30 estaban a favor de la propuesta. Suponga que p_M es la verdadera proporción subyacente de hombres que están a favor de la propuesta y que p_W es la verdadera proporción subyacente de mujeres que están a favor de la propuesta. Si en realidad es cierto que $p_W = p_M = p$, encuentre el MLE de la proporción común p.

***9.91** Encuentre el MLE de θ con base en una muestra aleatoria de tamaño n de una distribución uniforme en el intervalo $(0, 2\theta)$.

***9.92** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3}, & 0 \le y \le \theta, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el Ejercicio 9.52 se demostró que $Y_{(n)} = \text{máx}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

a Encuentre el MLE para θ. [*Sugerencia:* vea el Ejemplo 9.16.]

b Encuentre una función del MLE en el inciso a que sea una cantidad pivote. [*Sugerencia:* vea el Ejercicio 9.63.]

c Use la cantidad pivote del inciso b para determinar un intervalo de confianza de $100(1 - \alpha)\%$ para θ.

***9.93** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

En el ejercicio 9.53 se demostró que $Y_{(1)} = \text{mín}(Y_1, Y_2, \ldots, Y_n)$ es suficiente para θ.

a Encuentre el MLE para θ. [*Sugerencia*: véase el ejemplo 9.16.]

b Encuentre una función del MLE obtenido en el inciso a que sea una cantidad pivote.

c Utilice la cantidad pivote obtenida en el inciso b para encontrar un intervalo de confianza de 100 $(1-\alpha)\%$ para θ.

***9.94** Suponga que $\hat{\theta}$ es el MLE para un parámetro θ. Sea $t(\theta)$ una función de θ que posee una inversa única [es decir, si $\beta = t(\theta)$, entonces $\theta = t^{-1}(\beta)$]. Demuestre que $t(\hat{\theta})$ es el MLE de $t(\theta)$.

***9.95** Una muestra aleatoria de n piezas se selecciona de entre un número grande de piezas producidas por cierta línea de producción en un día. Encuentre el MLE de la relación R, la proporción de piezas defectuosas dividida entre la proporción de piezas buenas.

9.96 Considere una muestra aleatoria de tamaño n de una población normal con media μ y varianza σ^2, pero desconocida. Deduzca el MLE de σ.

9.97 La función de masa de probabilidad geométrica está dada por

$$p(y \mid p) = p(1-p)^{y-1}, \qquad y = 1, 2, 3, \ldots$$

Una muestra aleatoria de tamaño n se toma de una población con una distribución geométrica.

a Encuentre el estimador del método de momentos para p.

b Encuentre el estimador de máxima verosimilitud (MLE) para p.

9.8 Algunas propiedades de los estimadores de máxima verosimilitud con muestras grandes (opcional)

Los estimadores de máxima verosimilitud también tienen propiedades interesantes cuando se trabaja con muestras grandes. Suponga que $t(\theta)$ es una función derivable de θ. En la Sección 9.7 afirmamos por la propiedad de invarianza que si $\hat{\theta}$ es el MLE de θ, entonces el MLE de $t(\theta)$ está dado por $t(\hat{\theta})$. En algunas condiciones de regularidad que se cumplen para las distribuciones que consideraremos, $t(\hat{\theta})$ es un estimador *consistente* para $t(\theta)$. Además, para tamaños muestrales grandes,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\left[\dfrac{\partial t(\theta)}{\partial \theta}\right]^2 \Big/ nE\left[-\dfrac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]}}$$

tiene aproximadamente una distribución normal estándar. En esta expresión, la cantidad $f(Y \mid \theta)$ del denominador es la función de densidad correspondiente a la distribución continua de interés, evaluada en el valor aleatorio Y. En el caso discreto, el resultado análogo se cumple con la función de probabilidad evaluada en el valor aleatorio Y, $p(Y \mid \theta)$ se sustituye por la densidad $f(Y \mid \theta)$. Si deseamos un intervalo de confianza para $t(\theta)$, podemos usar a Z como la cantidad pivote. Si continuamos como en la Sección 8.6, obtenemos el siguiente intervalo de confianza aproximado de muestra grande 100(1 – α)% para $t(\theta)$:

$$t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]}$$

$$\approx t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left(\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]\right)\Bigg|_{\theta=\hat{\theta}}}.$$

Ilustramos esto con el siguiente ejemplo.

EJEMPLO **9.18** Para una variable aleatoria con distribución de Bernoulli, $p(y|p) = p^y(1-p)^{1-y}$, para $y = 0$, 1. Si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de esta distribución, deduzca un intervalo de confianza de 100(1 – α)% para $p(1 - p)$, la varianza asociada con esta distribución.

Solución Al igual que en el Ejemplo 9.14, el MLE del parámetro p está dado por $\hat{p} = W/n$ donde $W = \sum_{i=1}^{n} Y_i$. Se deduce que el MLE para $t(p) = p(1-p)$ es $\widehat{t(p)} = \hat{p}(1-\hat{p})$. En este caso,

$$t(p) = p(1-p) = p - p^2 \quad \text{y} \quad \frac{\partial t(p)}{\partial p} = 1 - 2p.$$

También,

$$\begin{aligned}
p(y \mid p) &= p^y(1-p)^{1-y}\\
\ln[p(y \mid p)] &= y(\ln p) + (1-y)\ln(1-p)\\
\frac{\partial \ln[p(y \mid p)]}{\partial p} &= \frac{y}{p} - \frac{1-y}{1-p}\\
\frac{\partial^2 \ln[p(y \mid p)]}{\partial p^2} &= -\frac{y}{p^2} - \frac{1-y}{(1-p)^2}\\
E\left\{-\frac{\partial^2 \ln[p(Y \mid p)]}{\partial p^2}\right\} &= E\left[\frac{Y}{p^2} + \frac{1-Y}{(1-p)^2}\right]\\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.
\end{aligned}$$

Sustituyendo en la fórmula anterior para el intervalo de confianza para $t(\theta)$, obtenemos

$$\begin{aligned}
&t(\hat{p}) \pm z_{\alpha/2}\sqrt{\left\{\left[\frac{\partial t(p)}{\partial p}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln p(Y \mid p)}{\partial p^2}\right]\right\}\Bigg|_{p=\hat{p}}}\\
&= \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\left\{(1-2p)^2 \Bigg/ n\left[\frac{1}{p(1-p)}\right]\right\}\Bigg|_{p=\hat{p}}}\\
&= \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})(1-2\hat{p})^2}{n}}
\end{aligned}$$

como el intervalo de confianza deseado para $p(1 - p)$. ■

Ejercicios

***9.98** Consulte el Ejercicio 9.97. ¿Cuál es la varianza aproximada del estimador de máxima verosimilitud?

***9.99** Considere la distribución estudiada en el Ejemplo 9.18. Use el método presentado en la Sección 9.8 para deducir un intervalo de confianza de $100(1 - \alpha)\%$ para $t(p) = p$. ¿El intervalo resultante le es conocido?

***9.100** Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de tamaño n de una distribución exponencial con media λ. Encuentre un intervalo de confianza de $100(1 - \alpha)\%$ para $t(\theta) = \theta^2$.

***9.101** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de tamaño n de una distribución de Poisson con media λ. Encuentre un intervalo de confianza de $100(1\text{-}\alpha)\%$ para $t(\lambda) = e^{-\lambda} = P(Y = 0)$.

***9.102** Consulte los Ejercicios 9.97 y 9.98. Si una muestra de tamaño 30 da $\overline{y} = 4.4$, encuentre un intervalo de confianza de 95% para p.

9.9 Resumen

En este capítulo continuamos y ampliamos el tema de estimación iniciado en el Capítulo 8. Los buenos estimadores son consistentes y eficientes cuando se comparan contra otros estimadores. Los estimadores más eficientes, los que tienen las varianzas más pequeñas, son funciones de los estadísticos suficientes que mejor resumen toda la información acerca del parámetro de interés.

Se presentaron dos métodos para determinar estimadores, el método de momentos y el método de máxima verosimilitud. Los estimadores de momento son consistentes pero por lo general no son muy eficientes. Los MLE, por otro lado, son consistentes y, si se ajustan para ser insesgados, con frecuencia llevan a estimadores insesgados de mínima varianza. Como tienen muy buenas propiedades, los MLE se usan con frecuencia en la práctica.

Bibliografía y lecturas adicionales

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

Cramer, H. 1973. *The Elements of Probability Theory and Some of Its Applications*, 2d ed. Huntington, N. Y.: Krieger.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.

Lindgren, B. W. 1993. *Statistical Theory*, 4th ed. Boca Raton, Fla.: Chapman and Hall/CRC.

Miller, I., and M. Miller. 2003. *John E. Freund's Mathematical Statistics with Applications*, 7th ed. Upper Saddle River, N. J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill.

Serfling, R. J. 2002. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Wilks, S. S. 1963. *Mathematical Statistics*. New York: Wiley.

Ejercicios complementarios

9.103 Una muestra aleatoria de tamaño n se toma de una población con una distribución de Rayleigh. Al igual que en el Ejercicio 9.34, la función de densidad de Rayleigh es

$$f(y) = \begin{cases} \left(\dfrac{2y}{\theta}\right)e^{-y^2/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre el estimador de probabilidad máxima de θ.

***b** Encuentre la varianza aproximada del estimador de máxima verosimilitud obtenido en el inciso a.

9.104 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de la función de densidad

$$f(y \mid \theta) = \begin{cases} e^{-(y-\theta)}, & y > \theta, \\ 0, & \text{en cualquier otro punto} \end{cases}$$

donde θ es una constante positiva desconocida.

a Encuentre un estimador $\hat{\theta}_1$ para θ por el método de momentos.

b Encuentre un estimador $\hat{\theta}_2$ para θ por el método de máxima verosimilitud.

c Ajuste $\hat{\theta}_1$ y $\hat{\theta}_2$ para que sean insesgados. Encuentre la eficiencia de la $\hat{\theta}_1$ ajustada con respecto a la $\hat{\theta}_2$ ajustada.

9.105 Consulte el Ejercicio 9.38(b). En las condiciones ahí señaladas, encuentre el MLE de σ^2.

***9.106** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una distribución de Poisson con media λ. Encuentre el estimador insesgado de varianza mínima (MVUE) de $P(Y_i = 0) = e^{-\lambda}$. [*Sugerencia:* haga uso del teorema de Rao–Blackwell.]

9.107 Suponga que se toma una muestra aleatoria de mediciones de vida útil, $Y_1, Y_2, \ldots, Y_n$, de componentes cuya duración tiene una distribución exponencial con media θ. Con frecuencia resulta de interés estimar

$$\overline{F}(t) = 1 - F(t) = e^{-t/\theta},$$

la *confiabilidad* en el tiempo t de dicho componente. Para cualquier valor fijo de t, encuentre el MLE de $\overline{F}(t)$.

***9.108** El MLE obtenido en el Ejercicio 9.107 es una función del estadístico suficiente mínimo para θ, pero no es insesgado. Use el teorema de Rao–Blackwell para determinar el MVUE de $e^{-t/\theta}$ mediante los siguientes pasos:

a Sea

$$V = \begin{cases} 1, & Y_1 > t, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Demuestre que V es un estimador insesgado de $e^{-t/\theta}$.

b Como $U = \sum_{i=1}^{n} Y_i$ es el estadístico suficiente mínimo para θ, demuestre que la función de densidad condicional para Y_1, dada $U = u$, es

$$f_{Y_1|U}(y_1 \mid u) = \begin{cases} \left(\dfrac{n-1}{u^{n-1}}\right)(u - y_1)^{n-2}, & 0 < y_1 < u, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

c Demuestre que

$$E(V \mid U) = P(Y_1 > t \mid U) = \left(1 - \frac{t}{U}\right)^{n-1}.$$

Éste es el MVUE de $t^{-t/\theta}$ por el teorema de Rao–Blackwell y por el hecho de que la función de densidad para U está completa.

***9.109** Suponga que n enteros se sacan al azar y *con reemplazo* de los enteros $1, 2, \ldots, N$. Esto es, cada entero muestreado tiene probabilidad $1/N$ de tomar cualquiera de los valores $1, 2, \ldots, N$ y los valores muestreados son independientes.

a Encuentre el estimador de método de momentos $\hat{N}_1$ de N.

b Encuentre $E(\hat{N}_1)$ y $V(\hat{N}_1)$.

***9.110** Consulte el Ejercicio 9.109.

a Encuentre el MLE $\hat{N}_2$ de N.

b Demuestre que $E(\hat{N}_2)$ es aproximadamente $[n/(n+1)]N$. Ajuste $\hat{N}_2$ para formar un estimador $\hat{N}_3$ que sea aproximadamente insesgado para N.

c Encuentre una varianza aproximada para $\hat{N}_3$ usando el hecho de que para N grande la varianza del máximo entero muestreado es aproximadamente

$$\frac{nN^2}{(n+1)^2(n+2)}.$$

d Demuestre que para N grande y $n > 1$, $V(\hat{N}_3) < V(\hat{N}_1)$.

***9.111** Consulte el Ejercicio 9.110. Suponga que tanques enemigos tienen números de serie $1, 2, \ldots, N$. Un espía observó al azar cinco tanques (con reemplazo) con números de serie 97, 64, 118, 210 y 57. Calcule N y ponga un límite en el error de estimación.

9.112 Sean $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Poisson con media λ y defina

$$W_n = \frac{\overline{Y} - \lambda}{\sqrt{\overline{Y}/n}}.$$

a Demuestre que la distribución de W_n converge en una distribución normal estándar.

b Utilice W_n y el resultado del inciso a para deducir la fórmula para un intervalo de confianza aproximado de 95% para λ.

CAPÍTULO 10

Prueba de hipótesis

10.1 Introducción

Recuerde que a menudo uno de los objetivos de la estadística es hacer inferencias acerca de parámetros poblacionales desconocidos con base en información contenida en datos muestrales. Estas inferencias se interpretan de dos formas: como estimaciones de los parámetros respectivos o como pruebas de hipótesis acerca de sus valores. Los Capítulos 8 y 9 se refieren a la estimación; en este capítulo examinamos el tema general de pruebas de hipótesis.

En muchos aspectos, el procedimiento formal para pruebas de hipótesis es semejante al método científico. Éste observa la naturaleza, formula una teoría y la confronta con lo observado. En nuestro contexto, el científico plantea una hipótesis respecto a uno o más parámetros poblacionales: de que son iguales a valores especificados. En seguida toma una muestra de la población y compara sus observaciones con la hipótesis. Si las observaciones no concuerdan con la hipótesis, las rechaza. De lo contrario, concluye que la hipótesis es verdadera o que la muestra no detectó la diferencia entre los valores real e hipotético de los parámetros poblacionales.

Por ejemplo, un investigador médico puede plantear la hipótesis de que un nuevo medicamento es más eficaz que otro para combatir una enfermedad. Para probarla, selecciona aleatoriamente pacientes infectados con la enfermedad y los divide en dos grupos. El nuevo medicamento A se administra a los pacientes del primer grupo y el B a los del segundo. Entonces, con base en el número de pacientes de cada grupo que se recuperen de la enfermedad, el investigador decide si el nuevo medicamento es más eficaz que el anterior.

Las pruebas de hipótesis se llevan a cabo en todos los campos en los que la teoría se pueda probar contra observación. Un ingeniero de control de calidad puede plantear la hipótesis de que un nuevo método de ensamble produce sólo 5% de piezas defectuosas. Un educador puede decir que dos métodos de enseñanza de lectura son igualmente eficaces, o un candidato político afirmar que la mayoría de los votantes está a favor de él. Todas estas hipótesis pueden ser tema de verificación estadística mediante el uso de datos muestrales observados.

¿Cuál es el papel de la estadística en pruebas de hipótesis? Dicho sin rodeos, ¿para qué sirve la estadística en este procedimiento de pruebas de hipótesis? Probar una hipótesis requiere tomar una decisión cuando se compara la muestra observada contra la teoría. ¿Cómo decidimos si la muestra no concuerda con la hipótesis del científico? ¿Cuándo debemos rechazar la hipótesis, cuándo debemos aceptarla y cuándo no revelar el juicio? ¿Cuál es la probabilidad de que tomemos una mala decisión y, en consecuencia, sufrir una pérdida? Y, en particular, ¿qué función de las mediciones muestrales debe emplearse para llegar a una decisión? Las respuestas a estas preguntas están contenidas en un estudio de pruebas de hipótesis estadísticas.

El Capítulo 8 introdujo el tema general de estimación y presentó algunos procedimientos de estimación intuitivos. El Capítulo 9 presentó unas propiedades de estimadores y métodos formales para obtener estimadores. Usamos el mismo método en nuestra exposición de prueba de hipótesis, es decir, introducimos el tema, presentamos procedimientos de prueba intuitivos y luego consideramos algunos métodos formales para deducir procedimientos de prueba de hipótesis estadística.

10.2 Elementos de una prueba estadística

Muchas veces, el objetivo de una prueba estadística es probar una hipótesis concerniente a los valores de uno o más parámetros poblacionales. Por lo general tenemos una teoría, es decir una *hipótesis de investigación*, acerca del o los parámetros que deseamos apoyar. Por ejemplo, suponga que un candidato, Jones, dice que él ganará más de 50% de los votos en una elección urbana y por tanto saldrá como ganador. Si no creemos en lo dicho por Jones, podríamos buscar apoyar la hipótesis de investigación de que Jones *no* está siendo favorecido por más de 50% del electorado. El apoyo para esta hipótesis de investigación, también llamada *hipótesis alternativa*, se obtiene mostrando (usando los datos muestrales como evidencia) que lo contrario de la hipótesis alternativa, llamado *hipótesis nula*, es falso. Entonces, una teoría se comprueba demostrando que no hay evidencia que sustente la teoría opuesta: en cierto sentido, una prueba por contradicción. Como buscamos apoyo para la hipótesis alternativa de que lo dicho por Jones es falso, nuestra hipótesis alternativa es que p, la probabilidad de seleccionar un votante que esté a favor de Jones, es menor que .5. Si podemos demostrar que los datos apoyan el rechazo de la hipótesis nula $p = .5$ (el valor mínimo necesario para una conseguir una mayoría) en favor de la hipótesis alternativa $p < .5$, hemos alcanzado nuestro objetivo de investigación. Aun cuando es común hablar de probar una hipótesis nula, el objetivo de investigación suele ser demostrar apoyo para la hipótesis alternativa, si dicho apoyo se justifica.

¿Cómo usamos esos datos para decidir entre la hipótesis nula y la hipótesis alternativa? Suponga que $n = 15$ votantes se seleccionan aleatoriamente de una ciudad y se registra Y, el número que está a favor de Jones. Si nadie en la muestra está a favor de Jones ($Y = 0$), ¿qué se concluiría acerca de lo dicho por Jones? Si Jones en realidad es favorecido por más de 50% del electorado, no es *imposible* observar que $Y = 0$ están a favor de Jones en una muestra de tamaño $n = 15$, pero es altamente *improbable*. Es mucho más probable que observemos $Y = 0$ si la hipótesis alternativa fuera cierta. Entonces, rechazaríamos la hipótesis nula ($p = .5$ a favor de la hipótesis alternativa ($p < .5$). Si observamos $Y = 1$ (o cualquier valor pequeño de Y), un razonamiento análogo nos lleva a la misma conclusión.

Cualquier prueba de hipótesis estadística funciona exactamente de la misma forma y está compuesta de los mismos elementos esenciales.

Los elementos de una prueba estadística

1. Hipótesis nula, H_0
2. Hipótesis alternativa, H_a
3. Estadístico de prueba
4. Región de rechazo

Para nuestro ejemplo, la hipótesis a ser probada, llamada *hipótesis nula* y denotada por H_0, es $p = .5$. La hipótesis alternativa (o investigación), denotada como H_a, es la hipótesis a ser aceptada en caso que H_0 sea rechazada. Por lo general la hipótesis alternativa es la que queremos comprobar con base en la información contenida en la muestra; así, en nuestro ejemplo, H_a es $p < .5$.

Las partes esenciales de una prueba estadística son el estadístico de prueba y una región de rechazo asociada. El *estadístico de prueba* (al igual que un estimador) es una función de las mediciones muestrales (Y en nuestro ejemplo) en las que la decisión estadística estará basada. La *región de rechazo*, que de aquí en adelante estará denotada por RR, especifica los valores del estadístico de prueba para el cual la hipótesis nula ha de ser *rechazada* a favor de la hipótesis alternativa. Si, para una muestra particular, el valor calculado del estadístico de prueba cae en la región de rechazo RR, rechazamos la hipótesis nula H_0 y aceptamos la hipótesis alternativa H_a. Si el valor del estadístico de prueba no cae en la RR, aceptamos H_0. Como ya indicamos antes, para nuestro ejemplo pequeños valores de Y nos llevarían a rechazar H_0. Por tanto, una región de rechazo que podríamos considerar es el conjunto de todos los valores de Y menores o iguales a 2. Usaremos la notación $\text{RR} = \{y : y \leq 2\}$, o bien, dicho de una forma más sencilla, $\text{RR} = \{y \leq 2\}$ para denotar esta región de rechazo.

Hallar una buena región de rechazo para una prueba estadística es un problema interesante que amerita más atención. Es evidente que pequeños valores de Y, por ejemplo $y \leq k$ (vea la Figura 10.1), son contradictorios para la hipótesis $H_0 : = .5$ pero favorables para la alternativa $H_a : p < .5$. Entonces de manera intuitiva seleccionamos la región de rechazo como $\text{RR} = \{y \leq k\}$. Pero, ¿qué valor debemos escoger para k? En forma más general, buscamos algunos

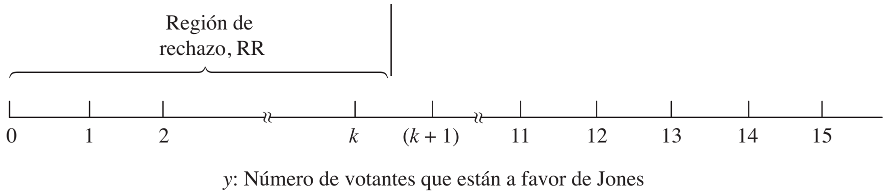


FIGURA **10.1** Región de rechazo, RR = $\{y \le k\}$, para una prueba de la hipótesis H_0: $p = .5$ contra la alternativa H_a: $p < .5$

criterios objetivos para decidir cuál valor de k especifica una buena región de rechazo de la forma $\{y \le k\}$.

Para cualquier región de rechazo fija (determinada por un valor particular de k), dos tipos de errores se pueden cometer al llegar a una decisión. Podemos decidir a favor de H_a cuando H_0 es verdadera (*error tipo I*), o podemos decidir a favor de H_0 cuando H_a es verdadera (*error tipo II*).

DEFINICIÓN **10.1**

Se comete un *error tipo I* si H_0 es rechazada cuando H_0 es verdadera. La *probabilidad de un error tipo I* está denotada por α. El valor de α se denomina *nivel* de la prueba.

Se comete un *error tipo II* si H_0 es aceptada cuando H_a es verdadera. La *probabilidad de un error tipo II* está denotada por β.

Para la encuesta política de Jones, cometer un error tipo I, es decir rechazar $H_0 : p = .5$ (y por tanto aceptar $H_a : p < .5$) cuando de hecho H_0 es verdadera, significa concluir que Jones perderá cuando en realidad va a ganar. En contraste, cometer un error tipo II significa aceptar $H_0 : p = .5$ cuando $p < .5$ y concluir que Jones ganará cuando en realidad va a perder. Para casi todas las situaciones reales, las decisiones incorrectas cuestan dinero, prestigio o tiempo e implican una pérdida. Entonces, α y β, las probabilidades de cometer estos dos tipos de error, miden los riesgos relacionados con las dos posibles decisiones erróneas que podrían resultar de una prueba estadística. Como tales, proporcionan una forma muy práctica de medir la bondad de una prueba.

EJEMPLO **10.1** Para la encuesta política de Jones se muestrearon $n = 15$ votantes. Deseamos probar $H_0 : p = .5$ contra la alternativa, $H_a : p < .5$. El estadístico de prueba es Y, el número de votantes muestreados a favor de Jones. Calcule α si seleccionamos RR $= \{y \le 2\}$ como la región de rechazo.

Solución Por definición,

$$\begin{aligned} \alpha &= P\,(\text{error tipo I}) = P\,(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera}) \\ &= P\,(\text{valor del estadístico de prueba está en RR cuando } H_0 \text{ es verdadera}) \\ &= P\,(Y \le 2 \text{ cuando } p = .5). \end{aligned}$$

Observe que Y es una variable aleatoria binomial con $n = 15$. Si H_0 es verdadera, $p = .5$ y obtenemos

$$\alpha = \sum_{y=0}^{2} \binom{15}{y} (.5)^y (.5)^{15-y} = \binom{15}{0}(.5)^{15} + \binom{15}{1}(.5)^{15} + \binom{15}{2}(.5)^{15}.$$

Usando la Tabla 1 del Apéndice 3 para evitar este cálculo, encontramos $\alpha = .004$. Entonces, si decidimos usar la región de rechazo RR = $\{y \le 2\}$, asumimos un riesgo muy pequeño ($\alpha =$ .004) de concluir que Jones perderá si en realidad es el ganador. ■

EJEMPLO **10.2** Consulte el Ejemplo 10.1. ¿Nuestra prueba es tan buena como para evitar concluir que Jones va a ganar si en realidad perderá? Suponga que él recibirá 30% de los votos ($p = .3$). ¿Cuál es la probabilidad β de que la muestra erróneamente nos lleve a concluir que H_0 es verdadera y que Jones v<a a ganar?

Solución Por definición,

$$\begin{aligned}\beta &= P\text{ (error tipo II)} = P\text{ (aceptar } H_0 \text{ cuando } H_a \text{ es verdadera)}\\ &= P\text{ (el valor del estadístico de prueba } \textit{no} \text{ está en RR cuando } H_a \text{ es verdadera).}\end{aligned}$$

Como buscamos calcular β cuando $p = .3$ (un valor particular de p que está en H_a),

$$\beta = P\,(Y > 2 \text{ cuando } p = .3) = \sum_{y=3}^{15} \binom{15}{y} (.3)^y (.7)^{15-y}.$$

De nuevo, consultando la Tabla 1, Apéndice 3, encontramos que $\beta = .873$. Si usamos RR = $\{y \le 2\}$, nuestra prueba por lo general nos llevará a concluir que Jones es ganador (con probabilidad $\beta = .873$), aun cuando p tan bajo como $p = .3$. ■

El valor de β depende del verdadero valor del parámetro p. Cuanto mayor sea la diferencia entre p y el valor hipotético (nulo) de $p = .5$, menor es la verosimilitud de que no rechacemos la hipótesis nula.

EJEMPLO **10.3** Consulte los Ejemplos 10.1 y 10.2. Calcule el valor de β si Jones recibirá sólo 10% de los votos ($p = .1$).

Solución En este caso, deseamos calcular β cuando $p = .1$ (otro valor particular de p en H_a).

$$\begin{aligned}\beta &= P\text{ (error tipo II)} = P\text{ (aceptar } H_0 \text{ cuando } p = .1)\\ &= P\text{ (el valor del estadístico de prueba } \textit{no} \text{ está en RR cuando } p = .1)\\ &= P\,(Y > 2 \text{ cuando } p = .1) = \sum_{y=3}^{15} \binom{15}{y} (.1)^y (.9)^{15-y} = .184.\end{aligned}$$

En consecuencia, si usamos $\{y \le 2\}$ como la región de rechazo, el valor de β cuando $p =$ .10 es menor que el valor para β que obtuvimos en el Ejemplo 10.2 con $p = .30$ (.184 contra .873). No obstante, cuando usamos esta región de rechazo, todavía tenemos una probabilidad bastante grande de decir que Jones es ganador si en verdad recibirá sólo 10% de los votos. ■

Los Ejemplos 10.1 al 10.3 muestran que la prueba usando RR $= \{y \le 2\}$ garantiza disminuir el riesgo de cometer un error tipo I ($\alpha = .004$), pero no ofrece protección adecuada contra un error tipo II. ¿Cómo podemos mejorar nuestra prueba? Una forma es balancear α y β al cambiar la región de rechazo. Si agrandamos RR en una nueva región de rechazo RR* (esto es, RR $\subset$ RR*), la prueba usando RR* nos llevará a rechazar H_0 con más frecuencia. Si α^* y α denotan las probabilidades de errores tipo I (niveles de las pruebas) cuando usamos RR* y RR como las regiones de rechazo, respectivamente, entonces, como RR $\subset$ RR*,

$$\begin{aligned}\alpha^* &= P\,(\text{estadístico de prueba está en RR}^* \text{ cuando } H_0 \text{ es verdadera})\\ &\ge P\,(\text{estadístico de prueba está en RR cuando } H_0 \text{ es verdadera}) = \alpha.\end{aligned}$$

Del mismo modo, si usamos la región de rechazo agrandada RR*, el procedimiento de prueba nos llevará a aceptar H_0 con menos frecuencia. Si β^* y β denotan las probabilidades de errores tipo II para las pruebas que usan RR* y RR, respectivamente, entonces

$$\begin{aligned}\beta^* &= P\,(\text{estadístico de prueba } \textit{no} \text{ está en RR}^* \text{ cuando } H_a \text{ es verdadera})\\ &\le P\,(\text{estadístico de prueba } \textit{no} \text{ está en RR cuando } H_a \text{ es verdadera}) = \beta.\end{aligned}$$

En consecuencia, si cambiamos la región de rechazo para aumentar α, entonces β disminuirá. Del mismo modo, si el cambio en la región de rechazo resulta en una disminución en α, entonces β aumentará. Por tanto, α y β están relacionadas de manera inversa.

EJEMPLO **10.4** Consulte las prueba analizada en el Ejemplo 10.1. Ahora suponga que RR $= \{y \le 5\}$. Calcule el nivel α de la prueba y calcule β si $p = .3$. Compare los resultados con los valores obtenidos en los Ejemplos 10.1 y 10.2 (donde usamos RR $= \{y \le 2\}$).

Solución En este caso,

$$\begin{aligned}\alpha &= P\,(\text{estadístico de prueba está en RR cuando } H_0 \text{ es verdadera})\\ &= P\,(Y \le 5 \text{ cuando } p = .5) = \sum_{y=0}^{5} \binom{15}{y} (.5)^{15} = .151.\end{aligned}$$

Cuando $p = .3$,

$$\begin{aligned}\beta &= P\,(\text{estadístico de prueba no está en RR cuando } H_a \text{ es verdadera y } p = .3)\\ &= P\,(Y > 5 \text{ cuando } p = .3) = \sum_{y=6}^{15} \binom{15}{y} (.3)^y (.7)^{15-y} = .278.\end{aligned}$$

Una comparación de las α y β calculadas aquí con los resultados de los Ejemplos 10.1 y 10.2 muestra que ampliar la región de rechazo de RR $= \{y \le 2\}$ a RR$^* = \{y \le 5\}$ aumenta α y disminuye β (véase la Tabla 10.1). Por tanto, hemos alcanzado un mejor balance entre

Tabla 10.1 Comparación de α y β para dos regiones de rechazo diferentes

	RR	
Probabilidades de error	$\{y \le 2\}$	$\{y \le 5\}$
α	.004	.151
β cuando $p = .3$	.873	.278

los riesgos de cometer errores tipo I y tipo II, aunque α como β siguen siendo todavía muy grandes. *¿Cómo podemos reducirlas*? La respuesta es clara y lógica: debemos obtener más información sobre la verdadera naturaleza de la población al aumentar el tamaño muestral. Para casi todas las pruebas estadísticas, si α está fija en algún valor aceptablemente pequeño, β disminuye cuando el tamaño muestral aumenta. ■

En esta sección hemos definido los elementos esenciales de cualquier prueba estadística. Hemos visto que se pueden cometer dos posibles tipos de error al probar una hipótesis: los errores tipo I y tipo II. Las probabilidades de cometer estos errores sirven como criterios para evaluar un procedimiento de prueba. En las siguientes secciones usaremos las distribuciones muestrales que obtuvimos en el Capítulo 7 para desarrollar métodos que permitan probar hipótesis relacionadas con parámetros de frecuente interés práctico.

Ejercicios

10.1 Defina α y β para una prueba estadística de hipótesis.

10.2 Un investigador ha preparado un nivel de dosis de droga que según él, inducirá el sueño en 80% de las personas que sufren de insomnio. Después de examinar la dosis, pensamos que lo dicho por él respecto a la efectividad de la dosis es exagerado. En un intento por refutar su dicho, administramos la dosis prescrita a 20 personas que padecen de insomnio y observamos Y, el número de individuos a quienes la dosis induce el sueño. Deseamos probar la hipótesis H_0: $p = .8$ contra la alternativa, $H_a : p < .8$. Suponga que se usa la región de rechazo $\{y \leq 12\}$.

a De acuerdo con la información de este problema, ¿qué es un error tipo I?
b Encuentre α.
c Con base en la información de este problema, ¿qué es un error tipo II?
d Encuentre β cuando $p = .6$.
e Encuentre β cuando $p = .4$.

10.3 Consulte el Ejercicio 10.2.

a Defina la región de rechazo de la forma $\{y \leq c\}$ de modo que $\alpha \approx .01$.
b Para la región de rechazo del inciso a, encuentre β cuando $p = .6$.
c Para la región de rechazo del inciso a, encuentre β cuando $p = .4$.

10.4 Suponga que deseamos probar la hipótesis nula H_0 de que la proporción p de hojas de contabilidad con errores es igual a .05 contra la alternativa H_a de que la proporción es mayor que .05 usando el siguiente esquema. Se seleccionan al azar dos hojas de contabilidad. Si ninguna de ellas tiene errores, rechazamos H_0; si una o más contienen un error, vemos una tercera hoja. Si ésta no tiene errores, rechazamos H_0. En todos los otros casos aceptamos H_0.

a De acuerdo con la información de este problema, ¿qué es un error tipo I?
b ¿Cuál es el valor de α relacionado con esta prueba?
c Con base en la información de este problema, ¿qué es un error tipo II?
d Calcule $\beta = P$ (error tipo II) como una función de p.

10.5 Suponga que Y_1 y Y_2 son independientes y están distribuidas idénticamente con una distribución uniforme en el intervalo $(\theta, \theta + 1)$. Para probar $H_0 : \theta = 0$ contra $H_a : \theta > 0$, tenemos dos pruebas:

Prueba 1: rechazar H_0 si $Y_1 > .95$.
Prueba 2: rechazar H_0 si $Y_1 + Y_2 > c$.

Encuentre el valor de c para que la prueba 2 tenga el mismo valor para α que la prueba 1. [*Sugerencia:* en el Ejemplo 6.3 obtuvimos las funciones de densidad y de distribución de la suma de dos variables aleatorias independientes que están distribuidas uniformemente en el intervalo (0, 1).]

10.6 Nos interesa probar si una moneda está o no balanceada, con base en el número de caras Y en 36 tiros de la moneda. ($H_0 : p = .5$ contra $H_a : p \neq .5$). Si usamos la región de rechazo $|y - 18| \geq 4$, ¿cuál es

a el valor de α?
b el valor de β si $p = .7$?

10.7 **Verdadero o falso** Consulte el Ejercicio 10.6.

a El nivel de la prueba calculado en el Ejercicio 10.6(a) es la probabilidad de que H_0 sea verdadera.
b El valor de β calculado en el Ejercicio 10.6(b) es la probabilidad de que H_a sea verdadera.
c En el Ejercicio 10.6(b), β se calculó suponiendo que la hipótesis nula era falsa.
d Si β se calculó cuando $p = 0.55$, el valor sería más grande que el valor de β obtenido en el Ejercicio 10.6(b).
e La probabilidad de que la prueba equivocadamente rechace H_0 es β.
f Suponga que la región de rechazo (RR) se cambió a $|y - 18| \geq 2$.

i Esta RR llevaría a rechazar la hipótesis nula con más frecuencia que la RR empleada en el Ejercicio 10.6.
ii Si α se calculó usando esta nueva RR, el valor sería más grande que el valor obtenido en el Ejercicio 10.6(a).
iii Si β se calculó cuando $p = .7$ y usando esta nueva RR, el valor sería más grande que el valor obtenido en el Ejercicio 10.6(b).

***10.8** Una prueba clínica en dos etapas está planeada para probar $H_0 : p = .10$ contra $H_a : p > .10$, donde p es la proporción de pacientes que responden a un tratamiento y que fueron tratados según el protocolo. En la primera etapa, 15 pacientes se acumularon y trataron. Si 4 o más de los que responden se observan entre los (primeros) 15 pacientes, H_0 es rechazada, el estudio se termina y no se acumulan más pacientes. De otro modo, otros 15 pacientes se acumularán y tratarán en la segunda etapa. Si un total de 6 o más de los que responden se observan entre los 30 pacientes acumulados en las dos etapas (15 en la primera etapa y 15 más en la segunda etapa), entonces H_0 es rechazada. Por ejemplo, si 5 de los que responden se encuentran entre los pacientes de la primera etapa, H_0 es rechazada y el estudio se termina. No obstante, si 2 de los que responden se encuentran entre los pacientes de la primera etapa, se acumulan 15 pacientes de la segunda etapa y se identifican otros 4 o más de los que responden (para un total de 6 o más entre los 30), H_0 es rechazada y el estudio termina.[1]

a Utilice la tabla binomial para hallar el valor numérico de α para este procedimiento de prueba.
b Utilice la tabla binomial para determinar la probabilidad de rechazar la hipótesis nula cuando use esta región de rechazo si $p = .30$.
c Para la región de rechazo definida líneas antes, encuentre β si $p = .30$.

1. Los ejercicios precedidos por un asterisco son opcionales.

10.3 Pruebas comunes con muestras grandes

Suponga que deseamos probar un conjunto de hipótesis respecto a un parámetro θ con base en una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$. En esta sección desarrollaremos procedimientos de prueba de hipótesis que están basados en un estimador $\hat{\theta}$ que tiene una distribución muestral normal (aproximadamente) con media θ y error estándar $\sigma_{\hat{\theta}}$. Los estimadores de muestra grande del Capítulo 8 (Tabla 8.1), por ejemplo $\overline{Y}$ y $\hat{p}$, satisfacen estos requisitos. También los satisfacen los estimadores empleados para comparar dos medias poblacionales $(\mu_1 - \mu_2)$ y para la comparación de dos parámetros binomiales $(p_1 - p_2)$.

Si θ_0 es un valor específico de θ, podemos probar $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$. La Figura 10.2 contiene una gráfica que ilustra las distribuciones muestrales de $\hat{\theta}$ para varios valores de θ. Si $\hat{\theta}$ es cercana a θ_0, parece razonable aceptar H_0. Pero, si en realidad $\theta > \theta_0$, es más probable que $\hat{\theta}$ sea más grande. En consecuencia, valores grandes de $\hat{\theta}$ (valores mayores a θ_0 en una cantidad apropiada) favorecen el rechazo de $H_0 : \theta = \theta_0$ y una aceptación de $H_a : \theta > \theta_0$. Esto es, las hipótesis nula y alternativa, el estadístico de prueba y la región de rechazo son como sigue:

$H_0 : \theta = \theta_0$.
$H_a : \theta > \theta_0$.
Estadístico de prueba: $\hat{\theta}$.
Región de rechazo: $\text{RR} = \{\hat{\theta} > k\}$ para alguna selección de k.

El valor real de k en la región de rechazo RR se determina al fijar la probabilidad α de error tipo I (el nivel de la prueba) y escoger k de conformidad (vea Figura 10.3). Si H_0 es verdadera, $\hat{\theta}$ tiene una distribución aproximadamente normal con media θ_0 y error estándar $\sigma_{\hat{\theta}}$. Por tanto,

FIGURA **10.2** Distribuciones muestrales del estimador $\hat{\theta}$ para varios valores de θ

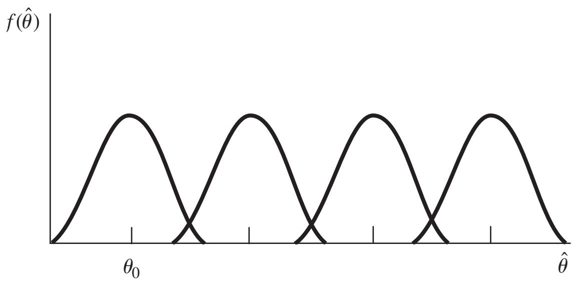
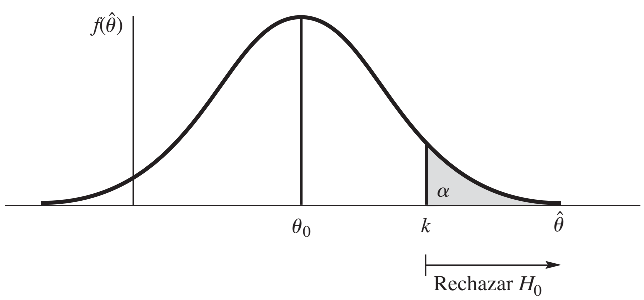


FIGURA **10.3** Región de rechazo de muestra grande para $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$



si deseamos una prueba de nivel α,

$$k = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$$

es la selección apropiada para k [si Z tiene una distribución normal estándar, entonces z_α es tal que $P(Z > z_\alpha) = \alpha$]. Como

$$\text{RR} = \{\hat{\theta} : \hat{\theta} > \theta_0 + z_\alpha \sigma_{\hat{\theta}}\} = \left\{\hat{\theta} : \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha\right\}$$

si $Z = (\hat{\theta} - \theta_0)/\sigma_{\hat{\theta}}$ se usa como estadístico de prueba, la región de rechazo también se puede escribir como RR $= \{z > z_\alpha\}$. Observe que Z mide el número de errores estándar entre el estimador para θ y θ_0, el valor de θ especificado en H_0. Por tanto, una forma equivalente de la prueba de hipótesis, con nivel α, es la siguiente:

$H_0 : \theta = \theta_0$.
$H_a : \theta > \theta_0$.
Estadístico de prueba: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$.

Región de rechazo: $\{z > z_\alpha\}$.

H_0 es rechazada si Z cae suficientemente alejada en la cola superior de la distribución normal estándar. La hipótesis alternativa $H_a : \theta > \theta_0$ se denomina alternativa de *cola superior* y RR $= \{z > z_\alpha\}$ se conoce como *región de rechazo de cola superior*. Observe que la fórmula precedente para Z es sencillamente

$$\text{Z} = \frac{\text{estimador para el parámetro– valor del parámetro dado por } H_0}{\text{error estándar del estimador}}$$

EJEMPLO **10.5** El vicepresidente de ventas de una gran empresa afirma que los vendedores están promediando no más de 15 contactos de venta por semana. (Le gustaría aumentar esta cantidad.) Como prueba de su afirmación, aleatoriamente se seleccionan $n = 36$ vendedores y se registra el número de contactos hechos por cada uno para una sola semana seleccionada al azar. La media y varianza de las 36 mediciones fueron 17 y 9, respectivamente. ¿La evidencia contradice lo dicho por el vicepresidente? Use una prueba con nivel $\alpha = .05$.

Solución Estamos interesados en la hipótesis de investigación de que la afirmación del vicepresidente es incorrecta. Esto se puede escribir formalmente como $H_a : \mu > 15$, donde μ es el número medio de contactos de ventas por semana. Por tanto, estamos interesados en probar

$$H_0 : \mu = 15 \quad \text{contra} \quad H_a : \mu > 15.$$

Sabemos que para n lo suficientemente grande, la media muestral $\overline{Y}$ es un estimador puntual de μ que está distribuido normalmente en forma aproximada con $\mu_{\overline{Y}} = \mu$ y $\sigma_{\overline{Y}} = \sigma/\sqrt{n}$. En consecuencia, nuestro estadístico de prueba es

$$Z = \frac{\overline{Y} - \mu_0}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu_0}{\sigma/\sqrt{n}}.$$

La región de rechazo, con $\alpha = .05$, está dada por $\{z > z_{.05} = 1.645\}$ (véase Tabla 4, Apéndice 3). La varianza poblacional σ^2 no se conoce, pero puede estimarse de manera muy precisa (porque $n = 36$ es lo suficientemente grande) con la varianza muestral $s^2 = 9$.

En consecuencia, el valor observado del estadístico de prueba es aproximadamente

$$z = \frac{\overline{y} - \mu}{s/\sqrt{n}} = \frac{17 - 15}{3/\sqrt{36}} = 4.$$

Debido a que el valor observado de Z se encuentra en la región de rechazo (porque $z = 4$ excede a $z_{.05} = 1.645$), rechazamos $H_0 : \mu = 15$. Entonces, al nivel de significancia $\alpha = .05$, la evidencia es suficiente para indicar que la afirmación del vicepresidente es incorrecta y que el número promedio de contactos de ventas por semana es mayor que 15. ■

EJEMPLO **10.6** Si la producción diaria de la máquina de una fábrica tiene más de 10% de artículos defectuosos, es necesario repararla. Una muestra aleatoria de 100 piezas de la producción del día contiene 15 piezas defectuosas y el supervisor decide que la máquina debe ser reparada. ¿La evidencia muestral apoya su decisión? Use una prueba con nivel .01.

Solución Si Y denota el número de piezas defectuosas observado, entonces Y es una variable aleatoria binomial, con p denotando la probabilidad de que una pieza seleccionada al azar sea defectuosa. En consecuencia, deseamos probar la hipótesis nula

$$H_0 : p = .10 \quad \text{contra la alternativa} \quad H_a : p > .10.$$

El estadístico de prueba, que está basado en $\hat{p} = Y/n$ (el estimador puntual insesgado de p), está dado por

$$Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}.$$

Podríamos haber usado $\sqrt{\hat{p}(1 - \hat{p})/n}$ para aproximar el error estándar de $\hat{p}$, pero como estamos considerando la distribución de Z bajo H_0, es más apropiado usar $\sqrt{p_0(1 - p_0)/n}$, el valor verdadero del error estándar de $\hat{p}$ cuando H_0 es verdadera.

De la Tabla 4, Apéndice 3, vemos que $P(Z > 2.33) = .01$ por lo cual tomamos $\{z > 2.33\}$ como la región de rechazo. El valor observado del estadístico de prueba está dado por

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{.15 - .10}{\sqrt{(.1)(.9)/100}} = \frac{5}{3} = 1.667.$$

Como el valor observado de Z no está en la región de rechazo, no podemos rechazar $H_0 : p = .10$ a favor de $H_a : p > .10$. En términos de su aplicación, concluimos que, en el nivel de significancia de $\alpha = .01$, la evidencia no apoya la decisión del supervisor.

¿Está equivocado el supervisor? No podemos hacer un juicio estadístico acerca de esto sino hasta que hayamos evaluado la probabilidad de aceptar H_0 cuando H_a es verdadera, es decir, hasta que hayamos calculado β. El método para calcular β se presenta en la Sección 10.4. ■

La prueba de $H_0 : \theta = \theta_0$ contra $H_a : \theta < \theta_0$ se hace de modo análogo, excepto que ahora rechazamos H_0 para valores de $\hat{\theta}$ que sean mucho menores que θ_0. El estadístico de prueba sigue siendo

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}},$$

pero para una α de nivel fijo rechazamos la hipótesis nula cuando $z < -z_\alpha$. Debido a que rechazamos H_0 a favor de H_a cuando z cae lo suficientemente lejos en la cola inferior de la distribución normal estándar, llamamos a $H_a : \theta < \theta_0$ una *cola inferior* alternativa y a RR: $\{z < -z_\alpha\}$ una *región de rechazo de cola inferior.*

Al probar $H_0 : \theta = \theta_0$ contra $H_a : \theta \neq \theta_0$, rechazamos H_0 si $\hat{\theta}$ es mucho menor o mucho mayor que θ_0. El estadístico de prueba es todavía Z, como antes, pero la región de rechazo está ubicada simétricamente en las dos colas de la distribución de probabilidad para Z. Entonces, rechazamos H_0 si $z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$. De un modo equivalente, rechazamos H_0 si $|z| > z_{\alpha/2}$. Esta prueba se llama *prueba de dos colas*, en contraposición a las *pruebas de una cola* empleadas para las alternativas $\theta < \theta_0$ y $\theta > \theta_0$. Las regiones de rechazo para la alternativa de cola inferior, $H_a : \theta < \theta_0$, y la alternativa de dos lados, $H_a : \theta \neq \theta_0$, se muestran en la Figura 10.4.

A continuación se proporciona un resumen de las pruebas de hipótesis de nivel α para muestras grandes desarrolladas hasta aquí.

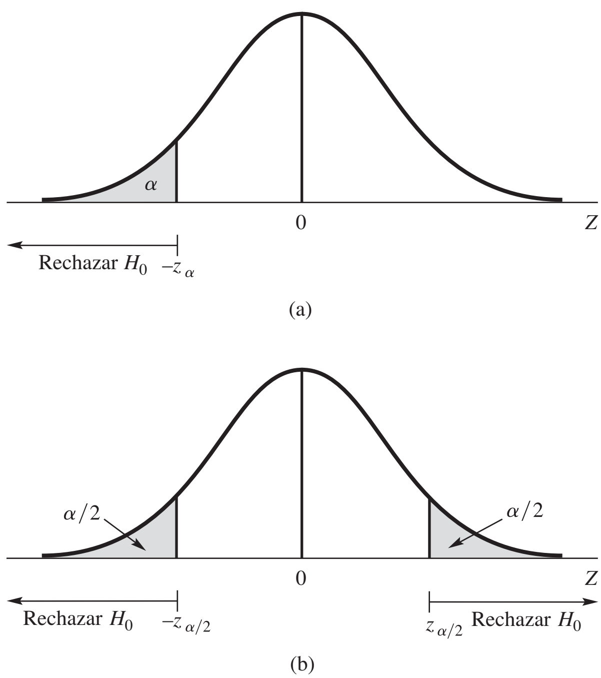


FIGURA **10.4** Regiones de rechazo para probar $H_0 : \theta = \theta_0$ contra (a) $H_a : \theta < \theta_0$ y (b) $H_a : \theta \neq \theta_0$, con base en $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$,

Pruebas de hipótesis de nivel α para muestras grandes

$$H_0 : \theta = \theta_0.$$

$$H_a : \begin{cases} \theta > \theta_0 & \text{(alternativa de cola superior).} \\ \theta < \theta_0 & \text{(alternativa de cola inferior).} \\ \theta \neq \theta_0 & \text{(alternativa de dos colas).} \end{cases}$$

$$\text{Estadístico de prueba: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}.$$

$$\text{Región de rechazo: } \begin{cases} \{z > z_\alpha\} & \text{(RR de cola superior)} \\ \{z < -z_\alpha\} & \text{(RR de cola inferior).} \\ \{|z| > z_{\alpha/2}\} & \text{(RR de dos colas).} \end{cases}$$

En cualquier prueba particular, sólo una de las alternativas citadas H_a es apropiada. Cualquiera que sea la hipótesis alternativa que escojamos, debemos estar seguros de usar la región de rechazo correspondiente.

¿Cómo decidir cuál hipótesis alternativa usar para una prueba? La respuesta depende de la hipótesis que pretendemos apoyar. Si estamos interesados sólo en detectar un aumento en el porcentaje de piezas defectuosas (Ejemplo 10.6), debemos localizar la región de rechazo en la cola superior de la distribución normal estándar. Por otra parte, si deseamos detectar un cambio en p ya sea arriba o debajo de $p = .10$, debemos localizar la región de rechazo en ambas colas de la distribución normal estándar y emplear una prueba de dos colas. El siguiente ejemplo ilustra una situación en la que una prueba de dos colas es apropiada.

EJEMPLO **10.7** Se realizó un estudio psicológico para comparar los tiempos de reacción de hombres y mujeres a un estímulo. En el experimento se emplearon muestras aleatorias independientes de 50 hombres y 50 mujeres. Los resultados se muestran en la Tabla 10.2. ¿Los datos presentan evidencia para sugerir una diferencia entre los tiempos medios de reacción verdaderos para hombres y mujeres? Use $\alpha = .05$.

Solución Con μ_1 y μ_2 denote los tiempos medios de reacción verdaderos para hombres y mujeres, respectivamente. Si deseamos probar la hipótesis de que las medias difieren, debemos probar $H_0 : (\mu_1 - \mu_2) = 0$ contra $H_a : (\mu_1 - \mu_2) \neq 0$. La alternativa de dos lados nos permite detectar ya sea el caso $\mu_1 > \mu_2$ o el caso inverso $\mu_2 > \mu_1$; en cualquier caso, H_0 es falsa.

El estimador puntual de $(\mu_1 - \mu_2)$ es $(\overline{Y}_1 - \overline{Y}_2)$. Como ya dijimos en las Secciones 8.3 y 8.6, debido a que las muestras son independientes y ambas son grandes, este estimador satisface las suposiciones necesarias para desarrollar una prueba de muestra grande. En conse-

Tabla 10.2 Datos para el Ejemplo 10.7

Hombres	Mujeres
$n_1 = 50$	$n_2 = 50$
$\overline{y}_1 = 3.6$ segundos	$\overline{y}_2 = 3.8$ segundos
$s_1^2 = .18$	$s_2^2 = .14$

cuencia, si deseamos probar $H_0 : \mu_1 - \mu_2 = D_0$ (donde D_0 es algún valor fijo) contra cualquier alternativa, el estadístico de prueba está dado por

$$Z = \frac{(\overline{Y}_1 - \overline{Y}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}},$$

donde σ_1^2 y σ_2^2 son las respectivas varianzas poblacionales. En esta aplicación deseamos usar una prueba de dos colas. Por tanto, para $\alpha = .05$, rechazamos H_0 para $|z| > z_{\alpha/2} = z_{.025} = 1.96$.

Para muestras grandes (por ejemplo $n_i > 30$), las varianzas muestrales dan buenas estimaciones de sus correspondientes varianzas poblacionales. Sustituyendo estos valores, junto con $\overline{y}_1, \overline{y}_2, n_1, n_2$ y $D_0 = 0$, en la fórmula para el estadístico de prueba, tenemos

$$z = \frac{\overline{y}_1 - \overline{y}_2 - 0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \approx \frac{3.6 - 3.8}{\sqrt{\dfrac{.18}{50} + \dfrac{.14}{50}}} = -2.5.$$

Este valor es menor que $-z_{\alpha/2} = -1.96$ y por tanto cae en la región de rechazo. En consecuencia, en el nivel $\alpha = .05$, concluimos que existe evidencia suficiente para permitirnos afirmar que los tiempos medios de reacción difieren para hombres y mujeres. ■

En esta sección, hemos descrito el procedimiento general para poner en práctica pruebas de hipótesis con muestras grandes para algunos parámetros de frecuente interés práctico. En la Sección 10.4 expondremos cómo calcular β, la probabilidad de un error tipo II, para estas pruebas con muestras grandes. La construcción de intervalos de confianza para estos parámetros y la implantación de pruebas formales de hipótesis son sorprendentemente semejantes. Ambos procedimientos usan los estimadores de los parámetros respectivos, los errores estándar de estos estimadores, así como cantidades obtenidas de la tabla de la distribución normal estándar. En la Sección 10.5 señalamos de manera explícita una correspondencia entre procedimientos de prueba con muestras grandes e intervalos de confianza con muestras grandes.

Ejercicios

10.9 **Ejercicio Applet** Use la aplicación *Hypothesis Testing (for Proportions)* para evaluar el impacto de cambiar el tamaño muestral en el valor de α. Cuando se tiene acceso a la aplicación, los valores predeterminados permitirán simulaciones, cuando el verdadero valor de $p = .5$, de pruebas Z repetidas de nivel $\alpha = .05$ para $H_0 : p = .5$ contra $H_a : p \neq .5$ y $n = 15$.

a ¿Qué acción califica como un "error" en la situación que será simulada?

b Haga clic en el botón "Draw Sample" para obtener los resultados relacionados con una sola muestra de tamaño 15. ¿Cuántos éxitos resultaron? ¿Cuál es el valor para $\hat{p}$? Calcule el valor de la prueba estadística de muestra grande. ¿Su cálculo está de acuerdo con el valor de z dado en la tabla bajo la curva normal? ¿El valor de z cae en la región de rechazo? ¿El resultado de esta simulación resultó en un error?

c Haga clic en el botón "Draw Sample" cinco veces más. ¿Cuántos valores diferentes para z observó usted? ¿Cuántos valores aparecieron en la región de rechazo dada por las colas de la curva normal?

d Haga clic en el botón “Draw Sample” hasta que obtenga una muestra simulada que resulte en el rechazo de H_0. ¿Cuál fue el valor de $\hat{p}$ que llevó al rechazo de H_0? ¿Cuántas pruebas realizó usted hasta que rechazó por primera vez H_0? ¿Por qué fueron necesarias tantas simulaciones hasta que rechazó por primera vez la nula?

e Haga clic en el botón “Draw 50 Samples” hasta que haya completado 200 o más simulaciones. Mueva el cursor sobre la caja sombreada arriba de “Reject” en la gráfica de barras de la parte inferior. ¿Qué proporción de las simulaciones resultó en el rechazo de H_0?

f ¿Por qué son exactamente de la misma altura las cajas arriba de “Reject” y “Error”?

g Use las flechas arriba y abajo a la derecha de la línea “n for sample” para cambiar el tamaño muestral para cada simulación a 20. Haga clic en el botón “Draw 50 Samples” hasta que haya simulado al menos 200 pruebas. ¿Qué proporción de las simulaciones resultó en el rechazo de H_0?

h Repita las instrucciones del inciso g para muestras de tamaño 30, 40 y 50. Haga clic en el botón “Show Summary” para ver los resultados de todas las simulaciones que realizó hasta aquí. ¿Qué observa acerca de las proporciones de veces que H_0 es rechazada usando muestras de tamaño 15, 20, 30, 40 y 50? ¿Le sorprenden estos resultados? ¿Por qué?

10.10 **Ejercicio Applet** Consulte el Ejercicio 10.9. Haga clic en el botón “Clear Summary” para borrar los resultados de cualesquiera simulaciones previas. Cambie el tamaño muestral de cada simulación a $n = 30$ y deje las hipótesis nula y alternativa en sus ajustes predeterminados $H_0 : p = .5$ y $H_a : p \neq .5$.

a Deje el verdadero valor de p en su ajuste predeterminado $p = .5$. Con esta situación, ¿qué es un error? Simule al menos 200 pruebas. ¿Qué proporción de las pruebas resultó en el rechazo de H_0? ¿Qué observa acerca de las alturas de las cajas arriba de “Reject” y “Error” en la gráfica inferior derecha? ¿Por qué?

b Deje sin cambio todos los ajustes, pero cambie el verdadero valor de p a .6. Con esta modificación, ¿qué es un error? Simule al menos 200 pruebas. ¿Qué proporción de las pruebas resultó en el rechazo de H_0? ¿Qué observa usted acerca de las alturas de las cajas arriba de “Reject” y “Error” en la gráfica inferior derecha? ¿Por qué?

c Deje sin cambio todos los ajustes del inciso b, pero cambie el verdadero valor de p a .7. Simule cuando menos 200 pruebas. Repita, ajustando el verdadero valor de p a .8. Haga clic en el botón “Show Summary”. Cuando el verdadero valor de p se aleje de .5 y se acerque a 1, ¿qué observa acerca de la proporción de simulaciones que llevan al rechazo de H_0? ¿Qué esperaría observar si se realiza un conjunto de simulaciones cuando el verdadero valor de p es .9?

d ¿Qué esperaría observar si se repitieran simulaciones cuando el valor real de p es .4, .3 y .2? Inténtelo.

10.11 **Ejercicio Applet** En el Ejercicio 10.9(h) observó que cuando la hipótesis nula es verdadera, para todos los tamaños muestrales la proporción de veces en que H_0 es rechazada es aproximadamente igual a α, la probabilidad de un error tipo I. Si probamos $H_0 : p = .5$, $H_a : p \neq .5$, ¿qué le ocurre al valor de β cuando aumenta el tamaño muestral? Fije el valor real de p en .6 y conserve el resto de los ajustes en sus valores predeterminados ($\alpha = .05$, $n = 15$).

a En la situación que será simulada, ¿cuál es la única clase de error que se puede cometer?

b Haga clic en el botón “Clear Summary”. Realice al menos 200 simulaciones. ¿Qué proporción de las simulaciones resultó en errores tipo II (mueva el cursor sobre la caja alrededor de “Error” en la parte inferior derecha de la pantalla)? ¿Cómo está relacionada la proporción de errores tipo II con la proporción de veces que H_0 es rechazada?

c Cambie n, el número de intentos usado para cada prueba simulada, a 30 y deje todos los otros ajustes sin cambio. Simule al menos 200 pruebas. Repita para $n = 50$ y $n = 100$. Haga clic en el botón “Show Summary”. ¿Cómo cambian los valores de $\beta(.6)$, la probabilidad de un error tipo II cuando $p = .6$, conforme aumenta el tamaño muestral?

d Deje la ventana con la información de resumen abierta y continúe con el Ejercicio 10.12.

10.12 **Ejercicio Applet** Consulte el Ejercicio 10.11. Cambie α a .1 pero conserve $H_0 : p = .5$, $H_a : p \neq .5$ y el verdadero valor de $p = .6$. Simule al menos 200 pruebas cuando $n = 15$. Repita para $n = 30$, 50 y 100. Haga clic en el botón "Show Summary". Ahora tendrá dos tablas de resumen (pudiera ser necesario arrastrar primero la última tabla de la parte superior). Compare los porcentajes de error cuando se simulen pruebas usando 15, 30, 50 y 100 intentos.

a ¿Cuál de las dos pruebas $\alpha = .05$ o $\alpha = .10$ proporciona los valores simulados más pequeños para β, usando muestras de tamaño 15?

b ¿Cuál proporciona los valores simulados más pequeños de β para cada uno de los otros tamaños muestrales?

10.13 **Ejercicio Applet** Si usted repitiera las instrucciones del Ejercicio 10.10, usando $n = 100$ en lugar de $n = 30$, ¿qué esperaría que fuera similar? ¿Qué esperaría que fuera diferente?

10.14 **Ejercicio Applet** Consulte el Ejercicio 10.9. Inicie la aplicación breve para probar $H_0 : p = .1$ contra $H_a : p < .1$ al hacer clic en el botón de radio "Lower" en la línea marcada "Tail" y ajustar el valor hipotético a .1. Fije el verdadero valor de $p = .1$, $n = .5$ y $\alpha = .20$.

a Haga clic en el botón "Draw Sample" hasta que obtenga una muestra con cero éxitos. ¿Cuál es el valor de z? ¿Cuál es el mínimo valor posible para z? ¿Es posible que obtenga una muestra de modo que el valor de z caiga en la región de rechazo? ¿Qué implica esto acerca de la probabilidad de que el procedimiento de prueba con "muestras grandes" rechace la hipótesis nula? ¿Este resultado invalida el uso de pruebas con muestras grandes para una proporción?

b ¿La prueba del inciso a rechazará la hipótesis nula verdadera aproximadamente 20% del tiempo si usamos $n = 10$? Inténtelo simulando al menos 100 pruebas. ¿Qué proporción de las simulaciones resulta en rechazo de la hipótesis nula?

c Examine los valores de $\hat{p}$ en la tabla bajo la curva normal e identifique el valor de $\hat{p}$ para el cual la hipótesis nula es rechazada. Use las tablas del apéndice para calcular la probabilidad de observar este valor cuando $n = 10$ y $p = .1$. ¿Este valor es cercano a .2?

d ¿Es $n = 100$ suficientemente grande para que la proporción simulada de rechazos sea cercana a .2? Simule al menos 100 pruebas y dé su respuesta con base en la simulación.

10.15 **Ejercicio Applet** Consulte el Ejercicio 10.10. Haga clic en el botón "Clear Summary" para borrar los resultados de cualquiera de las simulaciones previas. Cambie el tamaño muestral para cada simulación a $n = 30$ e inicie la aplicación breve para simular la prueba de $H_0 : p = .4$ contra $H_a : p > .4$ al nivel de significancia de .05.

a Haga clic en el botón "Clear Summary" para borrar los resultados o cualquiera de las simulaciones previas. Ajuste el valor real de p a .4 y haga al menos 200 simulaciones. ¿Cuál es el porcentaje de pruebas simuladas que resulta en el rechazo de la hipótesis nula? ¿La prueba funciona como se esperaba?

b Deje todos los ajustes como estaban en el inciso a pero cambie el valor real de p a .5. Simule al menos 200 pruebas. Repita cuando el valor real de p es .6 y .7. Haga clic en el botón "Show Summary". ¿Qué observa acerca del porcentaje de rechazo cuando el verdadero valor de p se aleja de .4 y se acerca a 1? ¿El patrón que se observa se asemeja a su idea de cómo debería operar una buena prueba?

10.16 **Ejercicio Applet** Consulte el Ejercicio 10.15. De nuevo, deseamos evaluar el desempeño de la prueba para $H_0 : p = .4$ contra $H_a : p > .4$ al nivel de significancia de .05 usando muestras de tamaño 30.

a Si el verdadero valor de p es .3, ¿aceptar la hipótesis alternativa es una decisión correcta o incorrecta?

b Haga clic en el botón "Clear Summary". Cambie el valor real de p a .3 y simule al menos 200 pruebas. ¿Qué fracción de las simulaciones resultó en la aceptación de la hipótesis alternativa?

c Cambie el valor real de p a .2 y simule al menos 200 pruebas. Haga clic en el botón "Show Summary". ¿Parece que algo está mal?

10.17 Un estudio publicado en la *American Journal of Sports Medicine*[2] reportó el número de metros (m) por semana nadados por dos grupos de nadadores —los que compitieron exclusivamente en brazada de pecho y los que compitieron en el relevo individual (que incluye brazada de pecho)—. Para cada nadador, se registró el número de metros por semana en la práctica de brazada de pecho y el resumen de estadísticas se proporciona a continuación. ¿Hay suficiente evidencia para indicar que el número promedio de metros por semana, empleados en practicar la brazada de pecho, es mayor para los especialistas en brazada de pecho y menor para los nadadores de relevo individual?

	Especialidad	
	Exclusivamente brazada de pecho	Relevo individual
Tamaño muestral	130	80
Media muestral (m)	9017	5853
Desviación muestral estándar (m)	7162	1961
Media poblacional	μ_1	μ_2

a Indique las hipótesis nula y alternativa.

b ¿Cuál es la región de rechazo apropiada para una prueba de nivel $\alpha = .01$?

c Calcule el valor observado del estadístico de prueba apropiado.

d ¿Cuál es su conclusión?

e ¿Cuál es la razón práctica para la conclusión a la que llegó en el inciso d?

10.18 Los salarios por hora en una industria particular están distribuidos normalmente con media de \$13.20 y desviación estándar de \$2.50. Una compañía en esta industria emplea 40 trabajadores, pagándoles un promedio de \$12.20 por hora. ¿Esta compañía puede ser acusada de pagar salarios abajo del estándar? Use una prueba de nivel $\alpha = .01$.

10.19 El voltaje de salida para un circuito eléctrico es de 130. Una muestra de 40 lecturas independientes del voltaje para este circuito dio una media muestral de 128.6 y desviación estándar de 2.1. Pruebe la hipótesis de que el promedio de voltaje de salida es 130 contra la alternativa de que es menor a 130. Use una prueba con nivel .05.

10.20 El índice Rockwell de dureza para acero se determina al presionar una punta de diamante en el acero y medir la profundidad de la penetración. Para 50 especímenes de una aleación de acero, el índice Rockwell de dureza promedió 62 con desviación estándar de 8. El fabricante dice que esta aleación tiene un índice de dureza promedio de al menos 64. ¿Hay suficiente evidencia para refutar lo dicho por el fabricante con un nivel de significancia de 1%?

10.21 Mediciones de resistencia al corte, obtenidas de pruebas de compresión no confinada para dos tipos de suelos, proporcionaron los resultados que se muestran en la siguiente tabla (medidas en toneladas por pie cuadrado). ¿Parecen diferir los suelos con respecto al promedio de resistencia al corte con un nivel de significancia de 1%?

Tipo de suelo I	Tipo de suelo II
$n_1 = 30$	$n_2 = 35$
$\overline{y}_1 = 1.65$	$\overline{y}_2 = 1.43$
$s_1 = 0.26$	$s_2 = 0.22$

2. *Fuente*: Kurt Grote, T. L. Lincoln y J. G. Gamble, "Hip Adductor Injury in Competitive Swimmers", *American Journal of Sports Medicine* 32(1)(2004): 104.

10.22 En el Ejercicio 8.66 examinamos los resultados de un estudio realizado en 2001 por Leonard, Speziale y Pernick que compara los métodos tradicionales y los orientados a actividades para enseñar biología. Se dieron resúmenes de estadísticas para las calificaciones de preexamen de 368 estudiantes a quienes posteriormente se les impartió clase usando el método tradicional y para 372 a quienes se les impartió clase usando el método orientado a actividades.

a Sin ver los datos, ¿esperaría usted que hubiera diferencia en las calificaciones medias de *preexamen* para quienes posteriormente se les impartió clase usando los métodos diferentes? Con base en su conjetura, ¿qué hipótesis alternativa escogería para probarla contra la hipótesis nula de que no hay diferencia en las calificaciones medias de preexamen para los dos grupos?

b ¿La hipótesis alternativa que usted propuso en el inciso a corresponde a una prueba estadística de una cola o de dos colas?

c La media y desviación estándar de las calificaciones de preexamen para quienes posteriormente se impartió clase usando el método tradicional fueron 14.06 y 5.45, respectivamente. Para quienes posteriormente se impartió clase usando el método orientado a actividades, las respectivas media y desviación estándar fueron 13.38 y 5.59. ¿Los datos dan apoyo a la conjetura de que las calificaciones medias de preexamen no difieren para estudiantes a quienes posteriormente se impartió clase usando los dos métodos? Pruebe con el uso de $\alpha = .01$.

10.23 Estudios realizados sobre los hábitos de venados de cola blanca indican que éstos viven y se alimentan en praderas muy limitadas, aproximadamente de 150 a 205 acres. Para determinar si difieren las praderas de venados situadas en dos zonas geográficas diferentes, los investigadores atraparon, marcaron y pusieron pequeños radiotransmisores a 40 venados. Varios meses después los venados fueron rastreados e identificados y se registró la distancia y desde el punto en que fueron soltados. La media y la desviación estándar de las distancias desde el punto en que fueron soltados se muestran en la siguiente tabla.[3]

	Ubicación	
	1	2
Tamaño muestral	40	40
Media muestral (ft)	2980	3205
Desviación muestral estándar (ft)	1140	963
Media poblacional	μ_1	μ_2

a Si usted no tiene una razón preconcebida para creer que una media poblacional es más grande que la otra, ¿qué seleccionaría para su hipótesis alternativa? ¿Y para su hipótesis nula?

b ¿Su hipótesis alternativa del inciso a implicaría una prueba de una cola o de dos colas? Explique.

c ¿Sus datos brindan suficiente evidencia para indicar que las distancias medias difieren para los dos lugares geográficos? Pruebe usando $\alpha = .10$.

10.24 Un estudio hecho por el Children's Hospital en Boston indica que alrededor de 67% de adultos estadounidenses y el 15% de niños y adolescentes tienen sobrepeso.[4] Trece niños de una muestra aleatoria de 100 se hallaron con sobrepeso. ¿Hay suficiente evidencia para indicar que el porcentaje publicado por el Children's Hospital es demasiado alto? Pruebe con un nivel de significancia de $\alpha = .05$.

10.25 Un artículo en *American Demographics* publica que 67% de adultos estadounidenses siempre vota en elecciones presidenciales.[5] Para probar esta afirmación, se tomó una muestra de 300 adultos y 192

3. *Fuente*: Charles Dickey, "A Strategy for Big Bucks", *Field and Stream*, octubre de 1990.

4. *Fuente*: Judy Holland, "'Cheeseburger Bill' on the Menu", *Press-Enterprise* (Riverside, Calif.), 9 de marzo de 2004, p. E1.

5. *Fuente*: Christopher Reynolds, "Rocking the Vote", *American Demographics,* febrero de 2004, p. 48.

dijeron que siempre votaban en elecciones presidenciales. ¿Los resultados de esta muestra proporcionan suficiente evidencia para indicar que el porcentaje de adultos que dicen que siempre votan en elecciones presidenciales, es diferente del porcentaje publicado en *American Demographics*? Pruebe usando $\alpha = .01$.

10.26 Según el *Washington Post*, casi 45% de todos los estadounidenses nacen con ojos cafés, aun cuando sus ojos no necesariamente continúan siendo cafés.[6] Una muestra aleatoria de 80 adultos encontró 32 con ojos cafés. ¿Hay suficiente evidencia en el nivel .01 para indicar que la proporción de adultos de ojos cafés difiere de la proporción de estadounidenses que nacen con ojos cafés?

10.27 El estado de California está trabajando muy duro para asegurar que todos los estudiantes en edad escolar, cuyo lenguaje materno no sea el inglés, tengan suficiente conocimiento de inglés cuando lleguen al sexto grado. Su avance es supervisado cada año por medio de un examen de desarrollo del idioma inglés en California. Los resultados para dos distritos del sur de California en el año escolar de 2003 se muestran en la siguiente tabla[7]. ¿Estos datos indican una diferencia importante en las proporciones de estudiantes del año escolar 2003 que dominan el inglés en los dos distritos? Use $\alpha = .01$.

Distrito	Riverside	Palm Springs
Número de estudiantes examinados	6124	5512
Porcentaje que domina el inglés	40	37

10.28 El mercantilismo del programa espacial de Estados Unidos ha sido un tema de gran interés desde que Dennis Tito pagó \$20 millones de dólares por viajar con cosmonautas rusos en un transbordador espacial.[8] En una encuesta hecha a 500 hombres y 500 mujeres, 20% de los hombres y 26% de las mujeres respondieron que el espacio debería permanecer libre de mercantilismo.

a ¿Existe evidencia, estadísticamente significativa, para sugerir que hay diferencia en las proporciones poblacionales de hombres y mujeres que piensan que el espacio debería permanecer libre de mercantilismo? Use una prueba de nivel .05.

b ¿Por qué una diferencia estadísticamente significativa en estas proporciones poblacionales es de importancia *práctica* para los anunciantes?

10.29 Un fabricante de lavadoras automáticas ofrece un modelo en uno de tres colores: A, B o C. De las primeras 1000 lavadoras vendidas, 400 eran del color A. ¿Se concluiría que los clientes tienen preferencia por el color A? Justifique su respuesta.

10.30 Una fabricante asegura que al menos 20% del público prefirió su producto. Se toma una muestra de 100 personas para comprobar esta afirmación. Con $\alpha = .05$, ¿qué tan pequeño necesitaría ser el porcentaje muestral antes de que la aseveración pueda ser refutada legítimamente? (Observe que esto requeriría una prueba de hipótesis de una cola.)

10.31 ¿Qué condiciones deben satisfacerse para que la prueba Z se utilice para comprobar una hipótesis respecto a una media poblacional μ?

10.32 En marzo de 2001, una encuesta de Gallup preguntó: "¿Cómo clasificaría usted la calidad general del medio ambiente en este país hoy en día: excelente, buena, regular o mala?". De 1060 adultos en todo el país, 46% dieron una clasificación de excelente o buena. ¿Es esto una evidencia convincente de que una mayoría de adultos de la nación piensa que la calidad del medio ambiente es regular o mala? Pruebe usando $\alpha = .05$.

6. *Fuente*: "Seeing the World Through Tinted Lenses", *Washington Post*, 16 de marzo de, 1993, p. 5.

7. *Fuente*: Cadonna Peyton, "Pupils Build English Skills", *Press-Enterprise* (Riverside, Calif.), 19 de marzo de, 2004, p. B-1.

8. *Fuente*: Adaptado de "Toplines: To de Moon?" *American Demographics*, agosto de 2001, p. 9.

10.33 Un politólogo cree que la fracción p_1 de republicanos es mayor que la fracción p_2 de demócratas que están a favor de la pena de muerte. Él adquirió muestras aleatorias independientes de 200 republicanos y 200 demócratas y encontró 46 republicanos y 34 demócratas a favor de la pena de muerte. ¿Esta evidencia proporciona apoyo estadístico para la creencia del investigador? Use $\alpha = .05$.

10.34 El Ejercicio 8.58 indicó que una muestra aleatoria de 500 mediciones de la permanencia en hospitales tenía una media muestral de 5.4 días y una desviación estándar muestral de 3.1 días. Un organismo federal de reglamentos plantea la hipótesis de que la permanencia tiene un exceso de 5 días. ¿Los datos apoyan esta hipótesis? Use $\alpha = .05$.

10.35 Michael Sosin[9] realizó un estudio reciente sobre los factores que explican por qué la gente que tiene casa pero se beneficia de los programas de alimentación, pierde su hogar. La siguiente tabla contiene los datos obtenidos en el estudio. ¿Hay suficiente evidencia para indicar que la proporción de los que actualmente trabajan es mayor para hombres con casa que para hombres sin casa? Use $\alpha = .01$.

	Hombres sin casa	Hombres con casa
Tamaño muestral	112	260
Número que actualmente trabaja	34	98

***10.36** Consulte el Ejercicio 8.68(b). ¿Hay evidencia de una diferencia entre la proporción de residentes que están a favor de la completa protección de lagartos y la proporción de quienes están a favor de la destrucción de los mismos? Use $\alpha = .01$.

10.4 Cálculo de las probabilidades del error tipo II y determinación del tamaño muestral para la prueba *Z*

El cálculo de β puede ser muy difícil para algunas pruebas estadísticas, pero es fácil hacerlo para las pruebas desarrolladas en la Sección 10.3. En consecuencia, podemos usar la prueba Z para demostrar tanto el cálculo de β como la lógica empleada para seleccionar el tamaño muestral para una prueba.

Para la prueba $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$, podemos calcular probabilidades de un error tipo II sólo para valores específicos de θ en H_a. Suponga que el experimentador tiene en mente una alternativa específica, por ejemplo $\theta = \theta_a$ (donde $\theta_a > \theta_0$). Como la región de rechazo es de la forma

$$\text{RR} = \{\hat{\theta} : \hat{\theta} > k\},$$

la probabilidad β de un error tipo II es

$$\begin{aligned}\beta &= P(\hat{\theta} \text{ no está en RR cuando } H_a \text{ es verdadera})\\ &= P(\hat{\theta} \le k \text{ cuando } \theta = \theta_a) = P\left(\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} \le \frac{k - \theta_a}{\sigma_{\hat{\theta}}} \text{ cuando } \theta = \theta_a\right).\end{aligned}$$

9. *Fuente*: Michael Sosin, "Homeless and Vulnerable Meal Program Users: A Comparison Study", *Social Problems* 39(2) (1992).

Si θ_a es el verdadero valor de θ, entonces $(\hat{\theta} - \theta_a)/\sigma_{\hat{\theta}}$ tiene una distribución normal estándar aproximadamente. En consecuencia, β puede ser determinada (en forma aproximada) al hallar un área correspondiente bajo una curva normal estándar.

Para una muestra fija de tamaño n, el tamaño de β depende de la distancia entre θ_a y θ_0. Si θ_a está cerca de θ_0, el verdadero valor de θ (ya sea θ_0 o θ_a) es difícil de detectar, y la probabilidad de aceptar H_0 cuando H_a es verdadera tiende a ser grande. Si θ_a está lejos de θ_0, el verdadero valor es relativamente fácil de detectar y β es mucho menor. Como vimos en la Sección 10.2, para un valor especificado de α, β puede hacerse menor si se escoge un tamaño muestral n grande.

EJEMPLO **10.8** Suponga que el vicepresidente del Ejemplo 10.5 desea detectar una diferencia igual a una llamada en el número medio de llamadas por semana. Es decir, desea probar $H_0 : \mu = 15$ contra $H_a : \mu = 16$. Con la información proporcionada en el Ejemplo 10.5, encuentre β para esta prueba.

Solución En el Ejemplo 10.5 teníamos $n = 36$, $\overline{y} = 17$ y $s^2 = 9$. La región de rechazo para una prueba de nivel .05 estaba dada por

$$z = \frac{\overline{y} - \mu_0}{\sigma/\sqrt{n}} > 1.645,$$

que es equivalente a

$$\overline{y} - \mu_0 > 1.645\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{o} \quad \overline{y} > \mu_0 + 1.645\left(\frac{\sigma}{\sqrt{n}}\right).$$

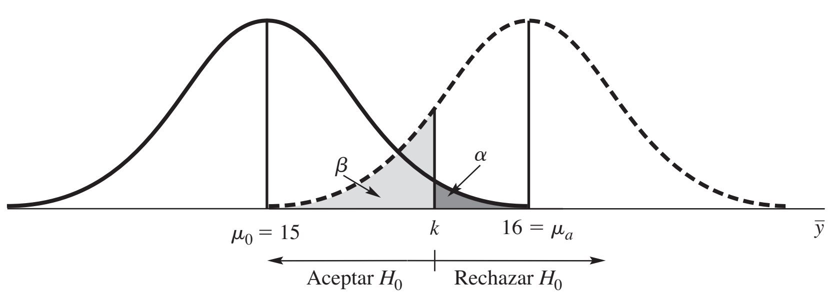
Sustituyendo $\mu_0 = 15$ y $n = 36$ y usando s para aproximar σ, encontramos que la región de rechazo es

$$\overline{y} > 15 + 1.645\left(\frac{3}{\sqrt{36}}\right), \text{ o bien, lo que es equivalente, } \overline{y} > 15.8225.$$

Esta región de rechazo se muestra en la Figura 10.5. Entonces, por definición, $\beta = P(\overline{Y} \leq 15.8225$ cuando $\mu = 16)$ está dada por el área sombreada bajo la curva de línea interrumpida a la izquierda de $k = 15.8225$ en la Figura 10.5. Por tanto, para $\mu_a = 16$,

$$\beta = P\left(\frac{\overline{Y} - \mu_a}{\sigma/\sqrt{n}} \leq \frac{15.8225 - 16}{3/\sqrt{36}}\right) = P(Z \leq -.36) = .3594$$

FIGURA **10.5** Región de rechazo para el Ejemplo 10.8 ($k = 15.8225$)



El valor grande de β nos dice que es frecuente que muestras de tamaño $n = 36$ no detecten una diferencia de 1 unidad de las medias hipotéticas. Podemos reducir el valor de β si aumentamos el tamaño muestral n. ■

El ejemplo anterior sugiere el procedimiento que un experimentador utiliza cuando escoge el o los tamaños muestrales para un experimento. Supongamos que usted desea probar $H_0 : \mu = \mu_0$ contra $H_a : \mu > \mu_0$. Si especifica los valores deseados de α y β (donde β se evalúa cuando $\mu = \mu_a$ y $\mu_a > \mu_0$), cualquier ajuste posterior de la prueba debe comprender dos cantidades restantes: el tamaño muestral n y el punto en el que empieza la región de rechazo, k. Como α y β se pueden escribir como probabilidades que contengan n y k, tenemos dos ecuaciones con dos incógnitas, de las que se puede despejar n simultáneamente. Entonces,

$$\begin{aligned}
\alpha &= P(\overline{Y} > k \text{ cuando } \mu = \mu_0) \\
&= P\left(\frac{\overline{Y} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu_0\right) = P(Z > z_\alpha), \\
\beta &= P(\overline{Y} \le k \text{ cuando } \mu = \mu_a) \\
&= P\left(\frac{\overline{Y} - \mu_a}{\sigma/\sqrt{n}} \le \frac{k - \mu_a}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu_a\right) = P(Z \le -z_\beta).
\end{aligned}$$

(Vea la Figura 10.5.)

De las ecuaciones previas para α y β, tenemos

$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = z_\alpha \quad \text{y} \quad \frac{k - \mu_a}{\sigma/\sqrt{n}} = -z_\beta.$$

Despejando k de ambas ecuaciones tendremos

$$k = \mu_0 + z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - z_\beta\left(\frac{\sigma}{\sqrt{n}}\right).$$

Por tanto,

$$(z_\alpha + z_\beta)\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_a - \mu_0, \text{ o bien, lo que es equivalente, } \sqrt{n} = \frac{(z_\alpha + z_\beta)\sigma}{(\mu_a - \mu_0)}.$$

Tamaño muestral para una prueba de cola superior de nivel α

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2}$$

Se obtendría exactamente la misma solución para una alternativa de una cola, $H_a : \mu = \mu_a$ con $\mu_a < \mu_0$. Se puede utilizar el método que acabamos de emplear para desarrollar una fórmula semejante para el tamaño muestral en un problema de prueba de hipótesis de una cola que satisfaga las condiciones de la Sección 10.3.

EJEMPLO **10.9** Suponga que el vicepresidente del Ejemplo 10.5 desea probar $H_0 : \mu = 15$ contra $H_a : \mu = 16$ con $\alpha = \beta = .05$. Determine el tamaño muestral que asegure esta precisión. Suponga que σ^2 es aproximadamente 9.

Solución Como $\alpha = \beta = .05$, se deduce que $z_\alpha = z_\beta = z_{.05} = 1.645$. Entonces

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_a - \mu_0)^2} = \frac{(1.645 + 1.645)^2(9)}{(16 - 15)^2} = 97.4.$$

En consecuencia, deben usarse $n = 98$ observaciones para satisfacer el requisito de que $\alpha \approx \beta \approx .05$ para la prueba del vicepresidente. ■

Ejercicios

10.37 Consulte el Ejercicio 10.19. Si el voltaje baja hasta 128 pueden aparecer consecuencias graves. Para probar $H_0 : \mu = 130$ contra $H_a : \mu = 128$, encuentre la probabilidad de un error tipo II, β, para la región de rechazo empleada en el Ejercicio 10.19.

10.38 Consulte el Ejercicio 10.20. El acero es suficientemente duro para satisfacer los requisitos de uso si la dureza media Rockwell no cae por debajo de 60. Usando la región de rechazo encontrada en el Ejercicio 10.20, encuentre β para la alternativa específica $\mu_a = 60$.

10.39 Consulte el Ejercicio 10.30. Calcule el valor de β para la alternativa $p_a = .15$.

10.40 Consulte el Ejercicio 10.33. El politólogo debería haber diseñado una prueba para la cual β es suficientemente pequeña cuando p_1 exceda a p_2 en una cantidad significativa. Por ejemplo, determine un tamaño muestral común n para una prueba con $\alpha = .05$ y $\beta \leq .20$ cuando en realidad p_1 excede a p_2 en .1. [*Sugerencia*: el valor máximo de $p(1 - p)$ es .25.]

10.41 Consulte el Ejercicio 10.34. Usando la región de rechazo ahí encontrada para calcular β cuando $\mu_a = 5.5$.

10.42 En los Ejercicios 10.34 y 10.41, ¿qué tan grande debe ser el tamaño muestral si requerimos que $\alpha = .01$ y $\beta = .05$ cuando $\mu_a = 5.5$?

10.43 Una muestra aleatoria de 37 estudiantes de segundo grado que practicaban deportes obtuvieron calificaciones de habilidad manual con una media de 32.19 y una desviación estándar de 4.34. Una muestra independiente de 37 estudiantes del mismo grado que no los practicaban tuvo calificaciones de destreza manual con media de 31.68 y desviación estándar de 4.56.

a Aplique una prueba para ver si existe suficiente evidencia que indique que los estudiantes de segundo grado que practican deportes tienen una calificación más alta en destreza manual. Use $\alpha = .05$.

b Para la región de rechazo empleada en el inciso a, calcule β cuando $\mu_1 - \mu_2 = 3$.

10.44 Consulte el Ejercicio 10.43. Encuentre los tamaños muestrales que dan como resultado $\alpha = .05$ y $\beta = .05$ cuando $\mu_1 - \mu_2 = 3$. (Suponga muestras de igual tamaño para cada grupo.)

10.5 Relaciones entre los procedimientos de pruebas de hipótesis e intervalos de confianza

Hasta este punto, hemos considerado dos procedimientos para muestras grandes que permiten hacer inferencias acerca de un parámetro objetivo θ. En la Sección 8.6 observamos que si $\hat{\theta}$ es un estimador de θ con una distribución muestral aproximadamente normal, un intervalo de confianza bilateral para θ con coeficiente de confianza $1 - \alpha$ está dado por

$$\hat{\theta} \pm z_{\alpha/2}\alpha_{\hat{\theta}}.$$

En esta expresión, $\sigma_{\hat{\theta}}$ es el error estándar del estimador $\hat{\theta}$ (la desviación estándar de la distribución muestral de $\hat{\theta}$) y $z_{\alpha/2}$ es un número obtenido usando la tabla normal estándar y tal que $P(Z > z_{\alpha/2}) = \alpha/2$. Para muestras grandes, si estuviéramos interesados en una prueba de nivel α de $H_0 : \theta = \theta_0$ contra la alternativa de dos lados $H_a : \theta \neq \theta_0$, los resultados de la sección anterior indican que usaríamos una prueba Z basada en el estadístico de prueba

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

y rechazaríamos H_0 si el valor de Z cayera en la región de rechazo $\{|z| > z_{\alpha/2}\}$. Estos dos procedimientos hacen fuerte uso del estimador $\hat{\theta}$, su error estándar $\sigma_{\hat{\theta}}$ y el valor de tabla $z_{\alpha/2}$. Exploremos estos dos procedimientos en forma más completa.

El complemento de la región de rechazo asociado con cualquier prueba recibe a veces el nombre de *región de aceptación* para la prueba. Para cualquiera de nuestras pruebas con nivel α de dos colas para muestras grandes, la región de aceptación está dada por $\overline{\text{RR}} = \{-z_{\alpha/2} \le z \le z_{\alpha/2}\}$. Esto es, no rechazamos $H_0 : \theta = \theta_0$ a favor de la alternativa de dos colas si

$$-z_{\alpha/2} \le \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \le z_{\alpha/2}.$$

Dicho de otro modo, la hipótesis nula no es rechazada (es "aceptada") en el nivel α si

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le \theta_0 \le \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}.$$

Observe que las cantidades de la extrema izquierda y extrema derecha de la secuencia previa de desigualdades son los puntos extremos izquierdo y derecho, respectivamente, de un intervalo de confianza $100(1 - \alpha)\%$ de dos lados para θ. Entonces, existe una dualidad entre nuestros procedimientos de muestra grande para construir un intervalo de confianza $100(1 - \alpha)\%$ de dos lados y para implantar una prueba de hipótesis de dos lados con nivel α. No rechace $H_0 : \theta = \theta_0$ a favor de $H_a : \theta \neq \theta_0$ si el valor θ_0 se encuentra dentro de un intervalo de confianza $100(1 - \alpha)\%$ para θ. Rechace H_0 si θ_0 se encuentra fuera del intervalo. De manera equivalente, un intervalo de confianza $100(1 - \alpha)\%$ de dos lados se puede interpretar como el conjunto de todos los valores de θ_0 para los cuales $H_0 : \theta = \theta_0$ es "aceptable" en el nivel α. Observe que *cualquier* valor dentro del intervalo de confianza es un valor aceptable del parámetro. No hay *un valor aceptable* para el parámetro sino muchos (en realidad, el número infinito de valores dentro del intervalo). Por esta razón, por lo general no *aceptamos* la hipótesis nula de que $\theta = \theta_0$, incluso si el valor θ_0 cae dentro de nuestro intervalo de confianza. Reconocemos que muchos valores de θ son aceptables y nos abstenemos de aceptar un valor individual de θ como *el* valor verdadero. Más comentarios respecto a la prueba de hipótesis se encuentran en la Sección 10.7.

Nuestra exposición anterior se concentró en la dualidad entre intervalos de confianza de dos lados y pruebas de hipótesis de dos lados. En los siguientes ejercicios de esta sección pediremos al lector que demuestre la correspondencia entre pruebas de hipótesis de muestra grande, pruebas de hipótesis unilaterales de nivel α y la construcción de los límites apropiados superior e inferior con coeficientes de confianza $1 - \alpha$. Si usted desea una prueba de nivel α de $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$ (una prueba de cola superior), debe aceptar la hipótesis alternativa si θ_0 es menor que el *límite de confianza inferior* $100(1 - \alpha)\%$ para θ. Si la hipótesis alternativa apropiada es $H_a : \theta < \theta_0$ (una prueba de cola inferior), usted debe rechazar $H_0 : \theta = \theta_0$ a favor de H_a si θ_0 es mayor que el *límite de confianza superior* $100(1 - \alpha)\%$ para θ.

Ejercicios

10.45 Consulte el Ejercicio 10.21. Construya un intervalo de confianza de 99% para la diferencia en resistencia media al corte para los dos tipos de suelo.

a ¿El valor $\mu_1 - \mu_2 = 0$ está dentro o fuera de este intervalo?

b Con base en el intervalo, ¿debe ser rechazada la hipótesis nula examinada en el Ejercicio 10.21? ¿Por qué?

c ¿Cómo se compara la conclusión a la que usted llegó con su conclusión en el Ejercicio 10.21?

10.46 Una prueba de hipótesis de nivel α y para muestras grandes en el caso de $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$ rechaza la hipótesis nula si

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_\alpha.$$

Demuestre que esto es equivalente a rechazar H_0 si θ_0 es menor que el límite de confianza inferior $100(1 - \alpha)\%$ de muestra grande para θ.

10.47 Consulte el Ejercicio 10.32. Construya un límite de confianza inferior de 95% para la proporción de adultos de la nación que piensan que la calidad del medio ambiente es regular o mala.

a ¿Cómo se compara el valor $p = .50$ con este límite inferior?

b Con base en el límite inferior del inciso a, ¿debe ser aceptada la hipótesis alternativa del Ejercicio 10.32?

c ¿Hay algún conflicto entre la respuesta del inciso b y su respuesta al Ejercicio 10.32?

10.48 Una prueba de hipótesis para muestras grandes y nivel α en el caso de $H_0 : \theta = \theta_0$ contra $H_a : \theta < \theta_0$ rechaza la hipótesis nula si

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} < -z_\alpha.$$

Demuestre que esto es equivalente a rechazar H_0 si θ_0 es mayor que el límite de confianza superior de muestra grande $100(1 - \alpha)\%$ para θ.

10.49 Consulte el Ejercicio 10.19. Construya un límite de confianza superior de 95% para la lectura promedio de voltaje.

a ¿Cómo se compara el valor $\mu = 130$ con este límite superior?

b Con base en el límite superior del inciso a, ¿debe aceptarse la hipótesis alternativa del Ejercicio 10.19?

c ¿Hay algún conflicto entre la respuesta del inciso b y su respuesta al Ejercicio 10.19?

10.6 Otra forma de presentar los resultados de una prueba estadística: niveles de significancia alcanzados o valores *p*

Como ya indicamos previamente, es frecuente que la probabilidad α de un error tipo I reciba el nombre de *nivel de significancia* o bien, dicho en forma más sencilla, *nivel* de la prueba. Aun cuando se recomienden con frecuencia pequeños valores de α, el valor real de α para usar en un análisis es un tanto arbitrario. Un experimentador puede escoger poner en práctica una prueba con $\alpha = .05$ mientras que otro podría preferir $\alpha = .01$. Es posible, por tanto, que dos personas analicen la misma información y lleguen a conclusiones opuestas, una que concluya que la hipótesis nula debe ser rechazada en el nivel de significancia $\alpha = .05$ y la otra que decida que la hipótesis nula debe ser rechazada con $\alpha = .01$. Además, los valores de α de .05 o .01 a menudo se emplean por costumbre o por comodidad más que como resultado de una cuidadosa consideración de las consecuencias de cometer un error tipo I.

Una vez tomada una decisión sobre un estadístico de prueba (Y en nuestro ejemplo de encuesta o una de las Z de la Sección 10.3), a veces es posible presentar el valor p o el nivel de significancia alcanzado y que está relacionado con una prueba. Esta cantidad es un estadístico que representa el valor más pequeño de α para el cual se puede rechazar la hipótesis nula.

DEFINICIÓN **10.2**

Si W es un estadístico de prueba, el *valor p*, o *nivel de significancia alcanzado*, es el nivel más pequeño de significancia α para el cual la información observada indica que la hipótesis nula debe ser rechazada.

Cuanto más pequeño sea el valor de p, es más fuerte la evidencia de que la hipótesis nula debe ser rechazada. Numerosas publicaciones científicas requieren que investigadores notifiquen los valores p relacionados con pruebas estadísticas, porque le dan al lector *más información* que la contenida en un informe de que la hipótesis nula fue o no rechazada por algún valor de α seleccionado por el investigador. Si el valor de p es lo suficientemente pequeño para convencer al lector, debe rechazar la hipótesis nula. Si un experimentador tiene un valor de α en mente, se puede usar el valor p para poner en práctica una prueba de nivel α. El valor p es el *más pequeño* de α para el cual la hipótesis nula puede ser rechazada. Entonces, si el valor deseado de α es mayor o igual al valor p, la hipótesis nula es rechazada para ese valor de α. En realidad, la hipótesis nula debería ser rechazada para cualquier valor de α *por abajo de p incluyendo* el valor de p. De otro modo, si α es menor que el valor p, la hipótesis nula no puede ser rechazada. En cierto sentido, el valor p permite que el lector de la investigación publicada evalúe la magnitud de la discrepancia entre los datos observados y la hipótesis nula.

En particular, el valor p permite a cada lector usar su propia elección para α para decidir si los datos observados conducen al rechazo de la hipótesis nula.

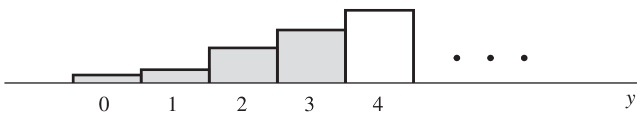
Los procedimientos para hallar valores p para las pruebas que hemos estudiado hasta aquí, se presentan en los siguientes ejemplos.

EJEMPLO **10.10** Recuerde el análisis de la encuesta política (vea Ejemplos 10.1 al 10.4) donde fueron muestreados $n = 15$ votantes. Si deseamos probar $H_0 : p = .5$ contra $H_a : p < .5$, usando $Y =$ número de votantes a favor de Jones como nuestro estadístico de prueba, ¿cuál es el valor p si $Y = 3$? Interprete el resultado.

Solución En exposiciones previas observamos que H_0 debería ser rechazada para valores pequeños de Y. En consecuencia, el valor p para esta prueba está dado por $P\{Y \leq 3\}$, donde Y tiene una distribución binomial con $n = 15$ y $p = .5$ (el área sombreada en la distribución binomial de la Figura 10.6). Usando la Tabla 1 del Apéndice 3, encontramos que el valor p es .018.

Como el valor $p = .018$ representa el valor más pequeño de α para el cual la hipótesis nula es rechazada, el experimentador que especifique cualquier valor de $\alpha \geq .018$ sería llevado a rechazar H_0 y concluir que Jones *no* tiene la mayoría del voto, pero si el experimentador escogió un valor α menor que .018 la hipótesis nula *no* podría ser rechazada.

FIGURA **10.6** Ilustración del valor p para el Ejemplo 10.10



■

Este ejemplo ilustra que reportar los valores p es particularmente benéfico cuando el estadístico de prueba apropiado posee una distribución *discreta.* En situaciones como ésta, es frecuente que no se pueda determinar alguna región de rechazo que dé un valor α de una magnitud especificada. Por ejemplo, en este caso no se puede determinar una región de rechazo de la forma $\{y \leq a\}$ para la cual $\alpha = .05$. En tales casos, reportar el valor p suele ser preferible a que el experimentador se limite a valores de α que se puedan obtener con base en la distribución discreta del estadístico de prueba.

El Ejemplo 10.10 también indica el método general para calcular valores p. Si fuéramos a rechazar H_0 en favor de H_a para valores pequeños de un estadístico de prueba W, por ejemplo RR : $\{w \leq k\}$, el valor p relacionado con un valor observado w_0 de W está dado por

$$\text{valor } p = P\,(W \leq w_0, \text{ cuando } H_0 \text{ es verdadera}).$$

Análogamente, si fuéramos a rechazar H_0 a favor de H_a para valores grandes de W, por ejemplo RR : $\{w \geq k\}$, el valor p relacionado con el valor observado w_0 es

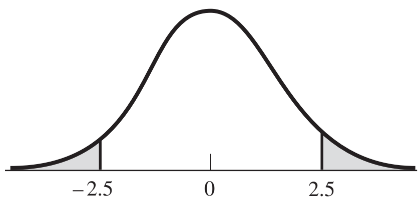
$$\text{valor } p = P\,(W \geq w_0, \text{ cuando } H_0 \text{ es verdadera}).$$

El cálculo de un valor p para una alternativa de dos colas se ilustra en el siguiente ejemplo.

EJEMPLO **10.11** Encuentre el valor p para el estadístico de prueba del Ejemplo 10.7.

Solución El Ejemplo 10.7 presenta una prueba de la hipótesis nula $H_0 : \mu_1 - \mu_2 = 0$ contra la hipótesis alternativa $H_a : \mu_1 - \mu_2 \neq 0$. El valor del estadístico de prueba, calculado a partir de los datos observados, fue $z = -2.5$. Como esta prueba es de dos colas, el valor p es la probabilidad de que $Z \leq -2.5$ o que $Z \geq 2.5$ (las áreas sombreadas en la Figura 10.7). De la Tabla 4, Apéndice 3, encontramos que $P(Z \geq 2.5) = P(Z \leq -2.5) = .0062$. Como se trata de una prueba de dos colas, el valor $p = 2(.0062) = .0124$. Por tanto, si $\alpha = .05$ (un valor mayor que .0124), rechazamos H_0 a favor de H_a y, de acuerdo con la conclusión del Ejemplo 10.7, deducimos que existe evidencia de una diferencia en el tiempo medio de reacción para hombres y mujeres. No obstante, si se escogiera $\alpha = .01$ (o cualquier valor de $\alpha < .0124$) *no* podríamos afirmar que hemos detectado una diferencia en los tiempos medios de reacción para los dos sexos.

FIGURA **10.7** Las áreas sombreadas dan el valor p para el Ejemplo 10.11.



■

Para las pruebas estadísticas que hemos desarrollado hasta ahora, el experimentador puede calcular valores p exactos con el uso de las tablas binomial y Z del Apéndice 3. La aplicación *Normal Probabilities* también se puede usar para calcular valores p relacionados con las pruebas Z estudiadas en las Secciones 10.3 y 10.4. Las tablas (del apéndice) de distribuciones para algunos de los estadísticos de prueba que encontraremos en secciones posteriores proporcionan valores críticos sólo para valores específicos de α (por ejemplo, .10, .05, .025, .01 y .005). En consecuencia, esas tablas no se pueden usar para calcular valores p exactos. No obstante, las tablas que se dan en el apéndice para las distribuciones F, t y χ^2 (y algunas otras) nos permiten determinar una región de valores dentro de las cuales se sabe que está el valor p. Por ejemplo, si el resultado de una prueba es estadísticamente significativo para $\alpha = .05$ pero no para $\alpha = .025$, indicaremos que $.025 \leq$ valor $p \leq .05$. Por tanto, para cualquier $\alpha \geq .05$, rechazamos la hipótesis nula; para $\alpha < .025$, *no* rechazamos la hipótesis nula; y para valores de α que estén entre .025 y .05, necesitamos buscar tablas más completas de la distribución apropiada antes de llegar a una conclusión. Las tablas del apéndice dan información útil acerca de valores p, pero los resultados son más bien engorrosos. Los valores p exactos relacionados con estadísticos de prueba con distribuciones t, χ^2 y F se obtienen fácilmente usando las aplicaciones que se presentan en el Capítulo 7. Muchas calculadoras también tienen capacidad de calcular valores p exactos.

La recomendación de que el investigador reporte el valor p para una prueba y deje la interpretación al lector no viola los procedimientos tradicionales de pruebas estadísticas (decisión teórica) descritos en las secciones anteriores. El informe de un valor p simplemente deja al lector decidir si rechaza la hipótesis nula (con el potencial relacionado de cometer errores tipo I o tipo II). Entonces, la responsabilidad de escoger α y, posiblemente, el problema de evaluar la probabilidad β de cometer un error tipo II se deja al lector.

Ejercicios

10.50 Los altos porcentajes de ocupación en vuelos regulares de líneas aéreas son esenciales para tener rentabilidad. Suponga que un vuelo regular debe promediar al menos 60% de ocupación para ser rentable y que un examen de los porcentajes de ocupación para 120 vuelos de las 10:00 de la mañana de Atlanta a Dallas mostraron un porcentaje medio de ocupación por vuelo de 58% y desviación estándar de 11%. Verifique si existe suficiente evidencia para apoyar la afirmación de que el vuelo no es rentable. Encuentre el valor *p* relacionado con la prueba. ¿Qué concluiría si desea poner en práctica la prueba en el nivel $\alpha = .10$?

10.51 A dos grupos de niños de escuela primaria se les enseñó a leer con el uso de métodos diferentes, 50 por cada método. Al término del periodo de instrucción, un examen de lectura arrojó los siguientes resultados $\overline{y}_1 = 74$, $\overline{y}_2 = 71$, $s_1 = 9$ y $s_2 = 10$.

a ¿Cuál es el nivel de significancia alcanzado si se desea verificar si la evidencia indica que hay una diferencia entre las dos medias poblacionales?

b ¿Qué se concluiría si desea un valor de α de .05?

10.52 Un biólogo ha lanzado la hipótesis de que una alta concentración de actinomicina D inhibe la síntesis del ácido ribonucleico en células y, por lo tanto, inhibe la producción de proteínas. Un experimento realizado para probar esta teoría comparó la síntesis del ácido ribonucleico en células tratadas con dos concentraciones de actinomicina D: 0.6 y 0.7 microgramos por litro. Las células tratadas con la concentración más baja (0.6) de actinomicina D indicaron que 55 de entre 70 se desarrollaron normalmente, mientras que sólo 23 de entre 70 parecieron desarrollarse normalmente con la concentración más alta (0.7). ¿Esta información indica que el porcentaje de síntesis normal de ácido ribonucleico es menor para células expuestas a las concentraciones más altas de actinomicina D?

a Encuentre el valor *p* para la prueba.

b Si usted elige usar $\alpha = .05$, ¿cuál es su conclusión?

10.53 ¿Cómo le gustaría llegar a vivir 200 años? Durante siglos la humanidad ha buscado la clave del misterio del envejecimiento. ¿Por qué envejecemos? ¿El envejecimiento puede hacerse lento? Los estudios se han enfocado en los *biomarcadores*, es decir, cambios físicos o biológicos que ocurren en un tiempo predecible en la vida de una persona. La teoría es que, si se encuentran formas para retardar la presencia de estos biomarcadores, la vida humana se puede prolongar. Un biomarcador clave, según científicos, es la capacidad vital forzada (FVC, por sus siglas en inglés), que es el volumen de aire que una persona puede exhalar después de hacer una profunda inhalación. Un estudio de 5209 hombres y mujeres de entre 30 y 62 años mostró que la FVC disminuyó, en promedio, 3.8 decilitros (dl) por década para hombres y 3.1 decilitros por década para mujeres.[10] Supongamos que usted desea determinar si un programa de entrenamiento físico, para hombres y mujeres de entre 50 y 60 años de edad, retardaría el envejecimiento; para hacerlo, se mide la FVC para 30 hombres y 30 mujeres que participan en el programa de entrenamiento a principios y fines del intervalo entre 50 a 60 años de edad y se registra el descenso de la FVC para cada persona. Un resumen de los datos aparece en la siguiente tabla.

	Hombres	Mujeres
Tamaño muestral	30	30
Promedio de descenso muestral en FVC (dl)	3.6	2.7
Desviación muestral estándar (dl)	1.1	1.2
Descenso en FVC en media poblacional	μ_1	μ_2

10. *Fuente*: T. Boddé, “Biomarkers of Aging: Key to a Younger Life”, *Bioscience* 31(8)(1981): 566−567.

a ¿Los datos aportan suficiente evidencia para indicar que la disminución de la FVC, en una década, para los hombres del programa de entrenamiento físico es menor que 3.8 dl? Encuentre el nivel de significancia alcanzado para la prueba.

b Consulte el inciso a. Si se escoge $\alpha = .05$, ¿los datos apoyan la controversia de que la disminución media en FVC es menor que 3.8 dl?

c Realice una prueba para determinar si la disminución en la FVC para mujeres en el programa de entrenamiento físico fue menor que 3.1 dl para una década. Encuentre el nivel de significancia alcanzado para la prueba.

d Consulte el inciso c. Si se escogió $\alpha = .05$, ¿los datos apoyan la controversia de que la disminución media en FVC es menor que 3.1 dl?

10.54 ¿Piensa usted que un porcentaje excepcionalmente alto de los ejecutivos de empresas grandes son diestros? Aun cuando 85% del público en general es diestro, un estudio a 300 principales ejecutivos de grandes empresas demostró que 96% eran diestros.

a ¿La diferencia en porcentajes es estadísticamente significativa? Pruebe usando $\alpha = .01$.

b Encuentre el valor p para la prueba y explique lo que significa.

10.55 Un servicio de cobro de cheques bancarios descubrió que alrededor de 5% de todos los cheques enviados al servicio no tenían fondos. Después de instituir un sistema de verificación de cheques para reducir sus pérdidas, el servicio encontró que de una muestra aleatoria de 1124 que fueron cobrados en efectivo sólo 45 cheques carecían de fondos. ¿Existe suficiente evidencia para afirmar que el sistema de verificación de cheques redujo la proporción de cheques sin fondos? ¿Qué nivel de significancia alcanzado está relacionado con la prueba? ¿Qué concluiría usted en el nivel $\alpha = .01$?

10.56 Una compañía farmacéutica realizó un experimento para comparar los tiempos medios, en días, necesarios para recuperarse de los efectos y complicaciones que siguen al inicio de un resfriado común. Este experimento comparó personas con dosis diaria de 500 miligramos (mg) de vitamina C con los que no recibieron un suplemento vitamínico. Para cada categoría de tratamiento se seleccionaron aleatoriamente 35 adultos y se encontró que los tiempos medios de recuperación y las desviaciones estándar para los dos grupos son los indicados en la siguiente tabla.

	Tratamiento	
	Sin suplemento	500 mg de vitamina C
Tamaño muestral	35	35
Media muestral	6.9	5.8
Desviación estándar muestral	2.9	1.2

a ¿Los datos indican que el uso de vitamina C reduce el tiempo medio requerido para la recuperación? Encuentre el nivel de significancia alcanzado.

b ¿Qué concluiría la compañía en el nivel $\alpha = .05$?

10.57 El editor de una revista ha encontrado que, según su experiencia, 60% de los suscriptores renuevan sus suscripciones. En una muestra aleatoria de 200 suscriptores, 108 indicaron que pensaban renovar sus suscripciones. ¿Cuál es el valor p relacionado con la prueba de que el porcentaje actual de renovaciones difiere del porcentaje antes experimentado?

10.58 En un estudio para evaluar diversos efectos de usar una modelo femenina para anunciar automóviles, a cada uno de 100 hombres se le mostraron fotografías de dos automóviles de mismo precio, color y tamaño pero de diferentes marcas. A 50 de los individuos (grupo A) se les mostró el automóvil 1 con una modelo femenina y el automóvil 2 sin modelo. Ambos automóviles se les mostraron sin la modelo a los otros 50 individuos (grupo B). En el grupo A, el automóvil 1 (mostrado con la modelo) fue calificado como más costoso por 37 individuos. En el grupo B, el automóvil 1 fue calificado como más costoso por 23 individuos. ¿Estos resultados indican que usar una modelo femenina aumenta el costo percibido de un automóvil? Encuentre el valor p relacionado e indique su conclusión para una prueba de nivel $\alpha = .05$.

10.7 Algunos comentarios respecto a la teoría de la prueba de hipótesis

Como ya señalamos antes, podemos elegir entre poner en práctica una prueba de una o de dos colas para una situación determinada. Esta elección está dictada por los aspectos prácticos del problema y depende del valor alternativo del parámetro θ que el experimentador está tratando de detectar. Si sufriéramos una gran pérdida financiera cuando θ fuera mayor que θ_0 pero no si fuera menor, concentraríamos nuestra atención en detectar valores de θ mayores que θ_0. En consecuencia, rechazaríamos la cola superior de la distribución para el estadístico de prueba previamente estudiado. Por otra parte, si estuviéramos igualmente interesados en detectar valores de θ menores o mayores que θ_0, emplearíamos una prueba de dos colas.

La teoría de pruebas estadísticas de las hipótesis, explicada en la Sección 10.2 y utilizada en la Sección 10.3, es un procedimiento bien definido que hace posible que el investigador rechace o acepte la hipótesis nula con riesgo medido α o β. Desafortunadamente, este sistema no es suficiente para todas las situaciones prácticas.

Para cualquier prueba estadística, la probabilidad α de un error tipo I depende del valor del parámetro especificado en la hipótesis nula. Esta probabilidad puede ser calculada, al menos aproximadamente, para cada uno de los procedimientos de prueba estudiados en este texto. Para los procedimientos que hemos visto hasta aquí, la probabilidad β de un error tipo II puede ser calculada sólo después que un valor *específico* del parámetro de interés haya sido seleccionado para considerarlo. La selección de un valor prácticamente significativo para este parámetro es, con frecuencia, difícil. Incluso si se puede identificar una alternativa significativa, el cálculo real de β es a veces bastante tedioso. La especificación de una hipótesis alternativa significativa es incluso más difícil para algunos de los procedimientos de prueba que consideraremos en los siguientes capítulos.

Desde luego, no deseamos pasar por alto la posibilidad de cometer un error tipo II. Más adelante en este capítulo determinaremos métodos para seleccionar pruebas con el más pequeño valor posible de β para pruebas donde α, la probabilidad de un error tipo I, es un valor fijo seleccionado por el investigador, pero incluso en estas situaciones el más pequeño valor posible de β puede ser bastante grande.

Estos obstáculos no impiden el uso de pruebas estadísticas; más bien, nos obligan a ser cuidadosos en el momento de sacar conclusiones donde hay evidencia insuficiente para permitir el rechazo de la hipótesis nula. Si se puede calcular un valor verdaderamente significativo para β, deberíamos sentirnos justificados al aceptar H_0 si el valor de β es pequeño y el valor del estadístico de prueba queda fuera de la región de rechazo. En una situación más típica, donde no hay un valor verdaderamente significativo para β, modificaremos nuestro procedimiento de la siguiente manera.

Cuando el valor del estadístico de prueba no se encuentre en la región de rechazo, “no rechazar” es mejor que “aceptar” la hipótesis nula. En el ejemplo de la encuesta que vimos en el Ejemplo 10.1, probamos $H_0 : p = .5$ contra $H_a : p < .5$. Si nuestro valor observado de Y

queda en la región de rechazo, rechazamos H_0 y decimos que la evidencia apoya la hipótesis de investigación de que Jones perderá. En esta situación, habremos demostrado que se apoya la hipótesis que nos interesa, es decir, la hipótesis de investigación. Sin embargo, si Y no cae en la región de rechazo y no podemos determinar un valor específico de p en H_a que sea de interés directo, simplemente decimos que *no* rechazamos H_0 y debemos buscar información adicional antes de llegar a una conclusión. De manera alternativa, podríamos indicar el valor de p relacionado con la prueba estadística y dejar la interpretación al lector.

Si H_0 es rechazada por un valor "pequeño" de α (o por un valor p pequeño), este suceso *no* implica que la hipótesis nula sea "errónea en una cantidad grande". *Sí* significa que la hipótesis nula puede ser rechazada con base en un procedimiento que incorrectamente rechaza la hipótesis nula (cuando H_0 es verdadera) con una pequeña probabilidad (esto es, con una pequeña probabilidad de un error tipo I). También debemos abstenernos de igualar la significancia *estadística* con la *práctica*. Si consideramos el experimento descrito y analizado en los Ejemplos 10.7 y 10.11, el valor p de .0124 es "pequeño", y el resultado es estadísticamente significativo para cualquier selección de $\alpha \geq .0124$. No obstante, la diferencia entre los tiempos medios de reacción para las dos muestras es de sólo .2 segundo, un resultado que puede o no ser *prácticamente* significativo. Para evaluar la significancia práctica de esa diferencia, es posible formar un intervalo de confianza para $\mu_1 - \mu_2$ con el uso de los métodos de la Sección 8.6.

Finalmente, son oportunos algunos comentarios respecto a la selección de las hipótesis nulas que hemos empleado, en particular en las pruebas de una cola. Así, en el Ejemplo 10.1 identificamos la hipótesis alternativa apropiada como $H_a : p < .5$ y empleamos $H_0 : p = .5$ como nuestra hipótesis nula. El estadístico de prueba fue Y = número de votantes que estaban a favor de Jones en una muestra de tamaño $n = 15$. Una región de rechazo que consideramos era $\{y \leq 2\}$. Usted podría preguntarse por qué no usamos $H_0^* : p \geq .5$ como la hipótesis nula. Esto tiene sentido porque *todo* posible valor de p está ya sea en $H_0^* : p \geq .5$ o en H_a: $p < 0.5$.

Entonces, ¿por qué usamos $H_0 : p = .5$? Sencillamente porque lo que en realidad nos importa es la *hipótesis alternativa* $H_a : p < .5$; la hipótesis nula no es nuestro principal interés. Como ya dijimos, por lo general no aceptamos en realidad la hipótesis nula de cualquier modo, cualquiera que sea su forma. Además, $H_0 : p = .5$ es más fácil de manejar y *lleva a exactamente las mismas conclusiones con el mismo valor de* α sin requerir que desarrollemos teoría adicional para manejar un $H_0^* : p \geq .5$. más complicado. Cuando usamos $H_0 : p = .5$ como hipótesis nula, calcular el nivel α de la prueba fue relativamente sencillo: sólo hallamos $P(Y \leq 2$ cuando $p = .5)$. Si hubiéramos usado $H_0^* : p \geq .5$ como la hipótesis nula, nuestra definición previa de α hubiera sido inadecuada porque el valor de $P(Y \leq 2)$ es en realidad una función de p para $p \geq .5$. En casos como éstos, α se define como el *máximo* (sobre todos los valores de $p \geq .5$) valor de $P(Y \leq 2)$. Aun cuando no deduzcamos este resultado aquí, $\text{máx}_{p \geq .5} P(Y \leq 2)$ ocurre cuando $p = .5$, el valor "frontera" de p en $H_0^* : p \geq .5$. Entonces, obtenemos el valor "correcto" de α si usamos la hipótesis nula más sencilla $H_0 : p = .5$.

Enunciados similares son verdaderos para todas las pruebas que hemos considerado hasta aquí y que tomaremos en cuenta para futuras explicaciones. Esto es, si consideramos que H_a: $\theta > \theta_0$ es la hipótesis de investigación adecuada, $\alpha = \text{máx}_{\theta \leq \theta_0} P$ (estadístico de prueba en RR) se presenta cuando $\theta = \theta_0$, el valor "frontera" de θ. Del mismo modo, si $H_a : \theta < \theta_0$ es la hipótesis de investigación adecuada, $\alpha = \text{máx}_{\theta \geq \theta_0} P$(estadístico de prueba en RR) se presenta cuando $\theta = \theta_0$. Por tanto, el uso de $H_0 : \theta = \theta_0$ en lugar de $H_0^* : \theta \geq \theta_0$ lleva al procedimiento de prueba correcto y el cálculo correcto de α sin hacer más consideraciones innecesarias.

Ejercicios

10.59 **Ejercicio Applet** Use la aplicación *Hypothesis Testing (for Proportions)* (consulte Ejercicios 10.9–10.16) para completar lo siguiente. Inicie la aplicación breve para simular los resultados de pruebas de $H_0 : p = .8$ contra $H_a : p > .8$, usando $\alpha = .2$ y muestras de tamaño $n = 30$. Haga clic en el botón "Clear Summary" para borrar los resultados de cualesquiera simulaciones previas.

a Fije el verdadero valor de p en .8 y realice al menos 200 pruebas simuladas. ¿Qué proporción de simulaciones resulta en el rechazo de la hipótesis nula?

b Deje todos los ajustes en sus valores previos, excepto que ahora cambie al verdadero valor de p a .75. Efectúe al menos 200 pruebas simuladas y observe la proporción de las simulaciones que llevaron al rechazo de la hipótesis nula. Repita, fijando el verdadero valor de p en .7 y de nuevo con el verdadero valor de $p = .65$.

c ¿Qué espera que ocurra si la simulación se repite después de fijar el verdadero valor de p en cualquier valor menor que .65? Inténtelo.

d Haga clic en el botón "Show Summary". ¿Cuál de las p verdaderas empleadas en las simulaciones resultó en la máxima proporción de prueba simulada que rechazó la nula y aceptó la alternativa, $H_a : p > .8$? ¿Esto confirma cualquier afirmación hecha en el último párrafo de la Sección 10.7? ¿Cuál afirmación?

10.60 **Ejercicio Applet** Consulte el Ejercicio 10.59. Inicie la aplicación breve para simular los resultados de pruebas de $H_0 : p = .4$ contra $H_a : p < .4$, usando $\alpha = .2$ y muestras de tamaño $n = 30$. Haga clic en el botón "Clear Summary" para borrar los resultados de cualesquiera simulaciones previas.

a Fije el verdadero valor de p en .4 y realice al menos 200 pruebas simuladas. ¿Qué proporción de las simulaciones resulta en el rechazo de la hipótesis nula?

b Deje todos los ajustes en sus valores previos, excepto que ahora cambie el verdadero valor de p a .45. Haga al menos 200 pruebas simuladas y observe la proporción de las simulaciones que llevaron al rechazo de la hipótesis nula. Repita, fijando el verdadero valor de p en .5, luego en .55.

c ¿Qué espera que ocurra si la simulación se repite después de fijar el verdadero valor de p en cualquier valor mayor que .55? Inténtelo.

d Haga clic en el botón "Show Summary". ¿Cuál de las p verdaderas empleadas en las simulaciones resultó en la máxima proporción de pruebas simuladas que rechazaron la nula y aceptaron la alternativa, $H_a : p < .4$? ¿Esto confirma cualesquiera afirmaciones hechas en el último párrafo de la Sección 10.7? ¿Cuáles de ellas?

10.8 Prueba de hipótesis con muestras pequeñas para μ y $\mu_1 - \mu_2$

En la Sección 10.3 explicamos los procedimientos de prueba de hipótesis que, al igual que los métodos de estimación de intervalo desarrollados en la Sección 8.6, son útiles para muestras grandes. Para que estos procedimientos sean aplicables, el tamaño muestral debe ser lo suficientemente grande para que $Z = (\hat{\theta} - \theta_0)/\sigma_{\hat{\theta}}$ tenga aproximadamente una distribución normal estándar. La Sección 8.8 contiene procedimientos basados en la distribución t para construir intervalos de confianza para μ (la media de una sola población normal) y $\mu_1 - \mu_2$ (la diferencia en las medias de dos poblaciones normales con varianzas iguales). En esta sección desarrollaremos procedimientos formales para probar hipótesis acerca de μ y $\mu_1 - \mu_2$, que son adecuados para muestras pequeñas a partir de poblaciones normales.

Supongamos que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de tamaño n de una distribución normal con media μ desconocida y varianza σ^2 desconocida. Si $\overline{Y}$ y S denotan la media muestral y la desviación muestral estándar, respectivamente, y si $H_0 : \mu = \mu_0$ es verdadera, entonces

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

tiene una distribución t con $n - 1$ grados de libertad (vea la Sección 8.8).

Como la distribución t es simétrica y en forma de campana, la región de rechazo para una prueba de la hipótesis $H_0 : \mu = \mu_0$ con muestras pequeñas debe estar localizada en las colas de la distribución t y ser determinada de forma semejante a la empleada con el estadístico Z de una muestra grande. Por analogía con la prueba Z desarrollada en la Sección 10.3, la región de rechazo adecuada para la alternativa de cola superior $H_a : \mu > \mu_0$ está dada por

$$\text{RR} = \{t > t_\alpha\},$$

donde t_α es tal que $P\{T > t_\alpha\} = \alpha$ para una distribución t con $n - 1$ grados de libertad (véase la Tabla 5, Apéndice 3).

Un resumen de las pruebas para μ con base en la distribución t, conocidas como *pruebas t*, es el siguiente.

Prueba de muestra pequeña para μ

Suposiciones: $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con $E(Y_i) = \mu$.

$H_0 : \mu = \mu_0$.

$$H_a : \begin{cases} \mu > \mu_0 & \text{(alternativa de cola superior).} \\ \mu < \mu_0 & \text{(alternativa de cola inferior).} \\ \mu \neq \mu_0 & \text{(alternativa de dos colas).} \end{cases}$$

Estadístico de prueba: $T = \dfrac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

$$\text{Región de rechazo:} \quad \begin{cases} t > t_\alpha & \text{(RR de cola superior).} \\ t < -t_\alpha & \text{(RR de cola inferior).} \\ |t| > t_{\alpha/2} & \text{(RR de dos colas).} \end{cases}$$

(Vea la Tabla 5, Apéndice 3, para valores de t_α, con $\nu = n - 1$ grados de libertad.)

EJEMPLO **10.12** El ejemplo 8.11 proporciona las velocidades iniciales de ocho balas probadas con una nueva pólvora, junto con la media muestral y la desviación muestral estándar, $\overline{y} = 2959$ y $s = 39.1$. El fabricante dice que la nueva pólvora produce un promedio de velocidad de no menos de 3000 pies por segundo. ¿Los datos muestrales aportan suficiente evidencia para contradecir lo afirmado por el fabricante en el nivel de significancia de .025?

Solución Suponiendo que las velocidades iniciales están distribuidas normalmente en forma aproximada, podemos usar la prueba que acabamos de mencionar. Deseamos probar $H_0 : \mu = 3000$ contra la alternativa, $H_a : \mu < 3000$. La región de rechazo está dada por $t < -t_{.025} = -2.365$, donde t posee $v = n - 1 = 7$ grados de libertad. Al hacer el cálculo encontramos que el valor observado del estadístico de prueba es

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} = \frac{2959 - 3000}{39.1/\sqrt{8}} = -2.966.$$

Este valor cae en la región de rechazo (esto es, $t = -2.966$ es menor que -2.365); por tanto, la hipótesis nula es rechazada en el nivel de significancia $\alpha = .025$. Concluimos que existe suficiente evidencia para contradecir lo afirmado por el fabricante y que la verdadera velocidad media es menor que 3000 ft por segundo en el nivel de significancia de .025. ■

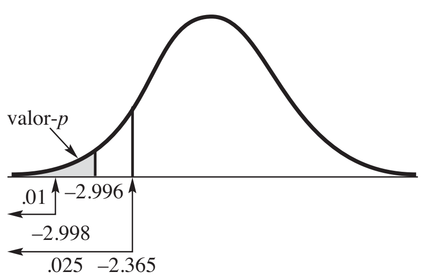
EJEMPLO **10.13** ¿Cuál es el valor p relacionado con el estadístico de prueba del Ejemplo 10.12?

Solución Como la hipótesis nula debe ser rechazada si t es "pequeña", el valor mínimo de α para el cual la hipótesis nula puede ser rechazada es elvalor $p = P\,(T < -2.966)$, donde T tiene una distribución t con $n - 1 = 7$ grados de libertad.

A diferencia de la tabla de áreas bajo la curva normal (Tabla 4, Apéndice 3), la Tabla 5 del Apéndice 3 no indica áreas correspondientes a muchos valores de t. En lugar de ello, proporciona los valores de t correspondientes a áreas de cola superior iguales a .10, .05, .025, .010 y .005. En vista de que la distribución t es simétrica alrededor de 0, podemos usar estas áreas de cola superior para dar las correspondientes áreas de cola inferior. En este caso, el estadístico t está basado en 7 grados de libertad; por tanto, consultamos la fila de grados de libertad (gl)= 7 de la Tabla 5 y encontramos que -2.966 cae entre $-t_{.025} = -2.365$ y $-t_{.01} = -2.998$. Estos valores están indicados en la Figura 10.8. Como el valor observado de $T\,(-2.966)$ es menor que $-t_{.025} = -2.376$ pero no menor que $-t_{.01} = -2.998$, rechazamos H_0 para $\alpha = .025$ pero no para $\alpha = .01$. Entonces, el valor p para la prueba satisface a $.01 \leq$ valor $p \leq .025$.

El valor exacto de p se obtiene fácilmente usando la aplicación *Student's t Probabilities and Quantiles* (disponible en www.thomsonedu.com/statistics/wackerly). Usando la aplicación con 7 grados de libertad, obtenemos un valor $p = P\,(T < -2.966) = P\,(T > 2.966) = .01046$, un valor que ciertamente está entre .01 y .025. Entonces, los datos indican que lo dicho por el fabricante debe ser rechazado para cualquier selección de $\alpha \geq .01046$.

FIGURA **10.8** Limitación del valor p para el Ejemplo 10.13 usando la Tabla 4, Apéndice 3



■

Una segunda aplicación de la distribución t es en la construcción de una prueba para muestra pequeña que compara las medias de dos poblaciones normales que poseen varianzas iguales. Suponga que se seleccionan muestras aleatorias independientes de cada una de dos poblaciones normales: $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ de la primera y $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ de la segunda, donde la media y la varianza de la i-ésima población son μ_i y σ^2, para $i = 1, 2$. Además, suponga que $\overline{Y}_i$ y S_i^2, para $i = 1, 2$, son las correspondientes medias muestrales y varianzas. Cuando estas suposiciones se cumplen, en la Sección 8.8 se demuestra que si

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

es el estimador agrupado para σ^2, entonces

$$T = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

tiene una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad. Si deseamos probar la hipótesis nula $H_0 : \mu_1 - \mu_2 = D_0$ para algún valor fijo D_0, se deduce que, si H_0 es verdadera, entonces

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

tiene una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad. Observe que este estadístico de prueba para muestras pequeñas es parecido a su similar para muestras grandes, el estadístico Z de la Sección 10.3. Las pruebas de la hipótesis $H_0 : \mu_1 - \mu_2 = D_0$ contra alternativas de cola superior, cola inferior y de dos colas son conducidas en la misma forma que en la prueba de muestras grandes, excepto que empleamos el estadístico t y tablas de la distribución t para llegar a nuestras conclusiones. Veamos a continuación un resumen de los procedimientos de prueba para muestras pequeñas en el caso de $\mu_1 - \mu_2$,

Pruebas con muestras pequeñas para comparar dos medias poblacionales

Suposiciones: muestras independientes de distribuciones normales con $\sigma_1^2 = \sigma_2^2$.

$H_0 : \mu_1 - \mu_2 = D_0$.

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0 & \text{(alternativa de cola superior).} \\ \mu_1 - \mu_2 < D_0 & \text{(alternativa de cola inferior).} \\ \mu_1 - \mu_2 \neq D_0 & \text{(alternativa de dos colas).} \end{cases}$$

Estadístico de prueba: $T = \dfrac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, donde $S_p = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$.

$$\text{Región de rechazo: } \begin{cases} t > t_\alpha & \text{(RR de cola superior).} \\ t < -t_\alpha & \text{(RR de cola inferior).} \\ |t| > t_{\alpha/2} & \text{(RR de dos colas).} \end{cases}$$

Aquí, $P(T > t_\alpha) = \alpha$ y grados de libertad $\nu = n_1 + n_2 - 2$. (Vea la Tabla 5, Apéndice 3.)

EJEMPLO **10.14** El Ejemplo 8.12 indica el tiempo requerido para completar un procedimiento de ensamble usando dos métodos diferentes de capacitación. Los datos muestrales son los que aparecen en la Tabla 10.3. ¿Hay suficiente evidencia para indicar una diferencia en los verdaderos tiempos medios de ensamble para quienes se capacitan usando los dos métodos? Pruebe al nivel $\alpha =$ .05 de significancia.

Tabla 10.3 Datos para el Ejemplo 10.14

Procedimiento estándar	Nuevo procedimiento
$n_1 = 9$	$n_2 = 9$
$\overline{y}_1 = 35.22$ segundos	$\overline{y}_2 = 31.56$ segundos
$\sum_{i=1}^{9}(y_{1i} - \overline{y}_1)^2 = 195.56$	$\sum_{i=1}^{9}(y_{2i} - \overline{y}_2)^2 = 160.22$

Solución Probaremos $H_0 : (\mu_1 - \mu_2) = 0$ contra la alternativa $H_a : (\mu_1 - \mu_2) \neq 0$. Por consiguiente, debemos usar una prueba de dos colas. El estadístico de prueba es

$$T = \frac{(\overline{Y}_1 - \overline{Y}_2) - D_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

con $D_0 = 0$, y la región de rechazo para $\alpha = .05$ es $|t| > t_{\alpha/2} = t_{.025}$. En este caso, $t_{.025} = 2.120$ porque t está basada en $(n_1 + n_2 - 2) = 9 + 9 - 2 = 16$ grados de libertad.

El valor observado del estadístico de prueba se encuentra al calcular primero

$$s_p = \sqrt{s_p^2} = \sqrt{\frac{195.56 + 160.22}{9 + 9 - 2}} = \sqrt{22.24} = 4.716.$$

Entonces,

$$t = \frac{\overline{y}_1 - \overline{y}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{35.22 - 31.56}{4.716\sqrt{\dfrac{1}{9} + \dfrac{1}{9}}} = 1.65.$$

Este valor no cae en la región de rechazo ($|t| > 2.120$); en consecuencia, la hipótesis nula no es rechazada. No hay suficiente evidencia para indicar una diferencia en los tiempos medios de ensamble para los dos periodos de capacitación en el nivel de significancia $\alpha = .05$.

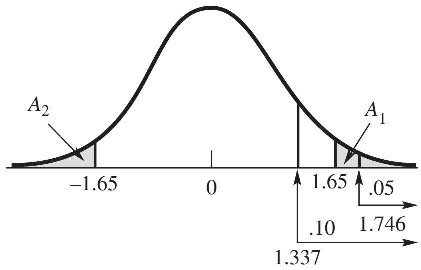
Observe que, de acuerdo con los comentarios de la Sección 10.7, no hemos aceptado H_0: $\mu_1 - \mu_2 = 0$ sino que, más bien, hemos expresado que no tenemos suficiente evidencia para rechazar H_0 y aceptar la alternativa $H_a : \mu_1 - \mu_2 \neq 0$. ■

EJEMPLO **10.15** Encuentre el valor p para la prueba estadística del Ejemplo 10.14.

Solución El valor observado del estadístico de prueba para esta prueba de dos colas fue $t = 1.65$. El valor p para esta prueba es entonces la probabilidad de que $T > 1.65$ o $T < -1.65$, las áreas sombreadas en la Figura 10.9, es decir, $A_1 + A_2$.

Como este estadístico de prueba está basado en $n_1 + n_2 - 2 = 16$ grados de libertad, consultamos la Tabla 5, Apéndice 3, para hallar $t_{.05} = 1.746$ y $t_{.10} = 1.337$. Por tanto, $A_1 = P(T > 1.65)$

FIGURA **10.9** Las áreas sombreadas son el valor de p para el Ejemplo 10.15



se encuentra entre .05 y .10; esto es, $.05 < A_1 < .1$. Del mismo modo, $.05 < A_2 < .1$. Como el valor $p = A_1 + A_2$, se deduce que $.1 <$ valor $p < .2$.

La aplicación *Student's t Probabilities and Quantiles* dice que, con 16 grados de libertad, $A_1 = P(T > 1.65) = .0592 = A_2$ y que el valor exacto de p es $= .1184$. Así, el valor mínimo de α para el cual los datos indican una diferencia en los tiempos medios de ensamble para los capacitados usando los dos métodos es .1184.

Ya sea que el valor de p se determine exactamente usando la aplicación o se limite usando la Tabla 5, Apéndice 3, si seleccionamos $\alpha = .05$, no podemos rechazar la hipótesis nula. Ésta es la misma conclusión a la que llegamos en el Ejemplo 10.14, donde formalmente pusimos en práctica la prueba de nivel de .05. ■

La prueba del Ejemplo 10.12 está basada en la suposición de que las mediciones de las velocidades iniciales han sido seleccionadas al azar de una población normal. En la mayoría de los casos es imposible verificar esta suposición. Podríamos preguntarnos cómo afecta este predicamento a la validez de nuestras conclusiones.

Estudios empíricos del estadístico de prueba

$$\frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

han sido realizados por muestreo de numerosas poblaciones con distribuciones no normales. Estas investigaciones han demostrado que las desviaciones moderadas de la normalidad en la distribución de la población tienen poco efecto en la distribución de probabilidad del estadístico de prueba. Este resultado, junto con la existencia común de distribuciones casi normales de datos en la naturaleza, hace que la prueba t de una media poblacional sea sumamente útil. Las pruebas estadísticas que no tienen sensibilidad a desviaciones a partir de las suposiciones en las que están basadas poseen una gran aplicabilidad. Debido a su insensibilidad a las violaciones de los supuestos formales, reciben el nombre de *pruebas estadísticas robustas*.

Al igual que la prueba t para una sola media, la prueba t para comparar dos medias poblacionales (a veces llamada *prueba t de dos muestras*) es robusta con respecto a la suposición de normalidad. También es robusta con respecto a la suposición de que $\sigma_1^2 = \sigma_2^2$ cuando n_1 y n_2 son iguales (o casi iguales).

Por último, la dualidad entre pruebas e intervalos de confianza que consideramos en la Sección 10.6 se cumple para las pruebas basadas en las distribuciones t que consideramos en esta sección y los intervalos de confianza presentados en la Sección 8.8.

Ejercicios

10.61 ¿Por qué la prueba Z suele ser inapropiada como procedimiento de prueba cuando el tamaño muestral es pequeño?

10.62 ¿Qué suposiciones se hacen cuando una prueba t de Student se emplea para probar una hipótesis que comprende una media poblacional?

10.63 Un proceso químico ha producido, en promedio, 800 toneladas de un químico al día. La producción diaria para la semana pasada fue 785, 805, 790, 793 y 802 toneladas.

a ¿Estos datos indican que el promedio de producción es menos de 800 toneladas y, por tanto, que algo está mal en el proceso? Pruebe al nivel de significancia de 5%. ¿Qué suposiciones deben cumplirse para que sea válido el procedimiento empleado por usted para analizar estos datos?

b Use la Tabla 5, Apéndice 3, para asignar límites para el valor p respectivo.

c **Ejercicio Applet** Use la aplicación *Student's t Probabilities and Quantiles* para hallar el valor p exacto. ¿El valor p exacto satisface los límites que obtuvo en el inciso b?

d Use el valor p del inciso c para decidir, en el nivel de significancia de 5%, si hay algo mal en el proceso. ¿La conclusión de usted concuerda con aquella a la que llegó en el inciso a?

10.64 Una máquina expendedora de gaseosas fue diseñada para descargar en promedio 7 onzas de líquido por taza. En una prueba de la máquina, diez tazas de líquido se sacaron de la máquina y se midieron. La media y la desviación estándar de las diez mediciones fueron 7.1 onzas y .12 onzas, respectivamente. ¿Estos datos presentan suficiente evidencia para indicar que la descarga media difiere de 7 onzas?

a ¿Qué se puede decir acerca del nivel de significancia alcanzado para esta prueba con base en la tabla t del apéndice?

b **Ejercicio Applet** Encuentre el valor exacto de p usando la aplicación *Student's t Probabilities and Quantiles*.

c ¿Cuál es la decisión adecuada si $\alpha = .10$?

10.65 Los operadores de vehículos a gasolina se quejan del precio de ésta en las gasolineras. Según el American Petroleum Institute, el impuesto federal a la gasolina por cada galón es constante (18.4¢ al 13 de enero de 2005), pero los impuestos estatales y locales varían de 7.5¢ a 32.10¢ para $n = 18$ áreas metropolitanas clave en todo el país.[11] El impuesto total por galón de gasolina en cada uno de estos lugares se da a continuación. Suponga que estas mediciones constituyen una muestra aleatoria de tamaño 18:

42.89	53.91	48.55	47.90	47.73	46.61
40.45	39.65	38.65	37.95	36.80	35.95
35.09	35.04	34.95	33.45	28.99	27.45

a ¿Hay suficiente evidencia para decir que el promedio de impuesto por galón de gasolina es menor que 45¢? Use la tabla t del apéndice para limitar el valor p relacionado con la prueba.

b **Ejercicio Applet** ¿Cuál es el valor p exacto?

c Construya un intervalo de confianza de 95% para el promedio de impuesto por galón de gasolina en Estados Unidos.

10.66 Los investigadores han demostrado que fumar cigarrillos tiene un efecto nocivo en la función de los pulmones. En su estudio del efecto de fumar sobre la capacidad de los pulmones para difundir el monóxido de carbono (DL, por sus siglas en inglés) Ronald Knudson, W. Kaltenborn y B. Burrows encontraron que

11. *Fuente*: "Gasoline Tax Rates by State", *http://www.gaspricewatch.com/usgastaxes.asp,* 13 de enero de 2005.

los fumadores actuales tuvieron lecturas de DL considerablemente más bajas que ex fumadores o que quienes no fuman.[12] La capacidad de difusión de monóxido de carbono para una muestra aleatoria de fumadores actuales fue como sigue:

103.768	88.602	73.003	123.086	91.052
92.295	61.675	90.677	84.023	76.014
100.615	88.017	71.210	82.115	89.222
102.754	108.579	73.154	106.755	90.479

¿Estos datos indican que la lectura media de DL para fumadores actuales es menor que 100, el promedio de lectura de DL para los no fumadores?

a Pruebe al nivel $\alpha = .01$.

b Dé límites al valor p usando una tabla del apéndice.

c **Ejercicio Applet** Encuentre el valor p exacto.

10.67 La información nutrimental dada por Kentucky Fried Chicken (KFC) dice que cada bolsa pequeña de papas contiene 4.8 onzas de alimento y 280 calorías. Una muestra de diez pedidos de restaurantes KFC en New York y New Jersey promedió 358 calorías.[13]

a Si la desviación estándar muestral fue $s = 54$, ¿hay suficiente evidencia para indicar que el número promedio de calorías en bolsas pequeñas de papas de KFC es mayor que el anunciado? Pruebe al nivel de significancia de 1%.

b Construya un límite inferior de confianza de 99% para el verdadero número medio de calorías en bolsas pequeñas de papas de KFC.

c Con base en el límite obtenido en el inciso b, ¿qué concluiría acerca de la afirmación de que el número medio de calorías es mayor que 280? ¿Cómo se compara esta conclusión con la del inciso a en la que se realizó una prueba formal de hipótesis?

10.68 ¿Qué supuestos se establecen acerca de las poblaciones de las que se obtienen muestras aleatorias independientes, cuando se usa la distribución t para hacer inferencias de muestras pequeñas respecto a las diferencias en medias poblacionales?

10.69 Dos métodos para enseñanza de lectura se aplicaron a dos grupos de niños de escuela primaria seleccionados al azar y luego se compararon con base en un examen de comprensión de lectura aplicado al final del periodo de enseñanza. Las medias muestrales y las varianzas calculadas a partir de las calificaciones del examen se muestran en la tabla siguiente.

	Método I	Método II
Número de niños por grupo	11	14
$\overline{y}$	64	69
s^2	52	71

¿Los datos presentan suficiente evidencia para indicar una diferencia en las calificaciones medias, para las poblaciones relacionadas con los dos métodos de enseñanza?

a ¿Qué se puede decir acerca del nivel de significancia alcanzado usando la tabla apropiada del apéndice?

12. *Fuente*: Ronald Knudson, W. Kaltenborn, and B. Burrows, "The Effects of Cigarette Smoking and Smoking Cessation on the Carbon Monoxide Diffussing Capacity of the Lung in Asymptomatic Subjects", *American Review of Respiratory Diseases* 140 (1989) 645−651.

13. *Fuente*: "KFC: Too Finger-Lickin' Good?", *Good Housekeeping* Saavy Consumer Product Tests, http://magazines.ivillage.com/goodhousekeeping/print/0,,446041,00.html, 11 de marzo de 2004.

b **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia alcanzado, usando la aplicación apropiada?

c ¿Qué suposiciones se requieren?

d ¿Qué concluiría usted en el nivel de significancia $\alpha = .05$?

10.70 La Florida Game and Fish Commission realizó un estudio para evaluar las cantidades de residuos químicos hallados en el tejido cerebral del pelícano café. En una prueba de DDT, muestras aleatorias de $n_1 = 10$ pelícanos jóvenes y $n_2 = 13$ polluelos produjeron los resultados que se muestran en la siguiente tabla (mediciones se expresan en partes por millón, ppm).

Jóvenes	Polluelos
$n_1 = 10$	$n_2 = 13$
$\overline{y}_1 = .041$	$\overline{y}_2 = .026$
$s_1 = .017$	$s_2 = .006$

a Pruebe la hipótesis de que cantidades medias de DDT, halladas en pelícanos jóvenes y polluelos, no difieren frente a la alternativa de que los pelícanos jóvenes tienen una media más grande. Use $\alpha = .05$ (Esta prueba tiene importantes implicaciones respecto a la acumulación de DDT con el tiempo.)

b ¿Hay evidencia de que la media para pelícanos jóvenes excede la de polluelos en más de .01 ppm?

i Limite el valor p usando una tabla del apéndice.

ii **Ejercicio Applet** Encuentre el valor exacto de p usando la aplicación apropiada.

10.71 En condiciones normales, ¿el promedio de temperatura corporal es igual para hombres y mujeres? Investigadores médicos interesados en esta pregunta recolectaron datos de un gran número de hombres y mujeres, y en la siguiente tabla se presentan muestras aleatorias de los datos.[14] ¿Hay suficiente evidencia para indicar que las temperaturas corporales medias difieren para hombres y mujeres?

Temperaturas corporales (ºF)	
Hombres	Mujeres
96.9	97.8
97.4	98.0
97.5	98.2
97.8	98.2
97.8	98.2
97.9	98.6
98.0	98.8
98.6	99.2
98.8	99.4

a Limite el valor de p, usando una tabla del apéndice.

b **Ejercicio Applet** Calcule el valor p.

10.72 Un artículo en *American Demographics* investigó los hábitos de consumo en un centro comercial. Tendemos a gastar más dinero en compras durante los fines de semana, en particular en domingos entre las 4:00 y las 6:00 p.m. Los compradores gastan menos los miércoles por la mañana.[15] Se seleccionaron

14. *Fuente*: *Journal of Statistics Education Data Archive,* http://www.amstat.org/publications/jse/jse-data-archive.html, marzo de 2006.

15. *Fuente*: John Fetto, "Shop Around the Clock", *American Demographics,* septiembre de 2003, p.18.

muestras aleatorias independientes de compradores y se registró la cantidad gastada por viaje al centro comercial, como se indica en la siguiente tabla:

Fines de semana	Días hábiles
$n_1 = 20$	$n_2 = 20$
$\overline{y}_1 = \$78$	$\overline{y}_2 = \$67$
$s_1 = \$22$	$s_2 = \$20$

a ¿Hay suficiente evidencia para decir que existe una diferencia en la cantidad promedio gastada por viaje en fines de semana y en días hábiles? Use $\alpha = .05$.

b ¿Cuál es el nivel de significancia alcanzado?

10.73 En el Ejercicio 8.83, presentamos algunos datos recolectados en un estudio hecho por Susan Beckham y sus colegas. En este estudio se hicieron mediciones de la presión del compartimento anterior del músculo (en milímetros de mercurio) para diez corredores sanos y diez ciclistas también sanos. El resumen de datos se repite aquí para comodidad del lector.

	Corredores		Ciclistas	
Condición	Media	s	Media	s
Reposo	14.5	3.92	11.1	3.98
80% consumo máx. de O_2	12.2	3.49	11.5	4.95

a ¿Hay suficiente evidencia para que se justifique decir que existe diferencia en las presiones medias del compartimento anterior del músculo para corredores y ciclistas que estén en reposo? Use $\alpha = .05$. Limite o determine el valor p respectivo.

b ¿Existe suficiente evidencia para permitirnos identificar una diferencia en las presiones medias del compartimento anterior del músculo para corredores y ciclistas al 80% de máximo consumo de O_2? Use $\alpha = .05$. Limite o determine el valor p respectivo.

10.74 Consulte el Ejercicio 8.88. Un informe de una prueba de laboratorio dice que, para estas especies de peces, el promedio de medición de LC50 es de 6 ppm. Use los datos del Ejercicio 8.88 para determinar si existe suficiente evidencia para indicar que el promedio de mediciones de LC50 es menor que 6 ppm. Use $\alpha = .05$.

10.75 El enorme crecimiento de la industria de la langosta de la Florida (llamada *langosta espinosa*), en los últimos 20 años, ha hecho que ésta sea la segunda industria de pesca más valiosa del estado. Se esperaba que una declaración del gobierno de las Bahamas, que prohibía a pescadores de Estados Unidos pescar en la parte de la plataforma continental de las Bahamas, redujera considerablemente la pesca en libras por trampa de langosta. De acuerdo con los registros, la pesca media anterior por trampa era de 30.31 libras. Un muestreo aleatorio de 20 trampas de langosta desde que la restricción de pesca impuesta por el gobierno de las Bahamas entró en vigor aportó los siguientes resultados (en libras):

17.4	18.9	39.6	34.4	19.6
33.7	37.2	43.4	41.7	27.5
24.1	39.6	12.2	25.5	22.1
29.3	21.1	23.8	43.2	24.4

¿Estos datos son suficiente evidencia para apoyar la afirmación de que la pesca media por trampa ha disminuido desde la imposición de las restricciones por el gobierno de las Bahamas? Pruebe usando $\alpha = .05$.

10.76 Jan Lindhe realizó un estudio[16] sobre el efecto de un enjuague bucal contra la acumulación de placa en los dientes. Catorce pacientes, cuyos dientes se limpiaron y pulieron muy bien, fueron asignados aleatoriamente a dos grupos de siete personas cada uno. A ambos grupos se les indicó usar enjuagues bucales (sin cepillarse) durante un periodo de 2 semanas. El grupo 1 utilizó un enjuague que contenía un agente contra la acumulación de placa; el grupo 2 de control recibió un enjuague bucal similar, excepto que, sin que los pacientes lo supieran, el enjuague no contenía el agente contra la acumulación de placa. El índice de placa y, que es una medida de la acumulación de placa, se registró a los 4, 7 y 14 días. La media y la desviación estándar para las mediciones de placa de 14 días para los dos grupos se dan en la siguiente tabla:

	Grupo de control	Grupo contra la placa
Tamaño muestral	7	7
Media	1.26	.78
Desviación estándar	.32	.32

a Exprese las hipótesis nula y alternativa que deberían usarse para probar la efectividad del enjuague bucal contra la placa.

b ¿Los datos dan suficiente evidencia para indicar que el enjuague bucal contra la placa es efectivo? Pruebe usando $\alpha = .05$.

c Limite o encuentre el valor p para la prueba.

10.77 En el Ejercicio 8.90 presentamos un resumen de datos respecto a calificaciones del Examen de Evaluación Escolar (SAT, por sus siglas en inglés) verbal y de matemáticas para estudiantes de preparatoria que pretendían una especialización en ingeniería o en idiomas y literatura. Los datos se resumen en la tabla siguiente:

Especialización probable	Verbal		Matemáticas	
Ingeniería ($n = 15$)	$\overline{y} = 446$	$s = 42$	$\overline{y} = 548$	$s = 57$
Idiomas/literatura ($n = 15$)	$\overline{y} = 534$	$s = 45$	$\overline{y} = 517$	$s = 52$

a ¿Hay suficiente evidencia para indicar una diferencia en las calificaciones medias verbales del SAT para estudiantes de preparatoria que pretenden una especialización en ingeniería y en idiomas/ literatura? Limite o determine el valor p respectivo. ¿Qué concluiría usted en el nivel de significancia de $\alpha = .05$?

b ¿Los resultados obtenidos en el inciso a son consistentes con los obtenidos en el Ejercicio 8.90(a)?

c Conteste las preguntas planteadas en el inciso a con relación a las calificaciones medias de matemáticas del SAT para los dos grupos de estudiantes.

d ¿Los resultados obtenidos en el inciso c son consistentes con los obtenidos en el Ejercicio 8.90(b)?

10.9 Pruebas de hipótesis referentes a varianzas

De nuevo suponga que tenemos una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de una distribución normal con media μ desconocida y varianza σ^2 desconocida. En la Sección 8.9 empleamos el método de pivote para construir un intervalo de confianza para el parámetro σ^2. En esta sección analizaremos el problema de probar $H_0 : \sigma^2 = \sigma_0^2$ para algún valor fijo σ_0^2 contra varias

16. *Fuente*: Jan Lindhe, "Clinical Assessment of Antiplaque Agents", *Compendium of Continuing Education in Dentistry,* supl. no. 5, 1984.

hipótesis alternativas. Si H_0 es verdadera y $\sigma^2 = \sigma_0^2$, el Teorema 7.3 implica que

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

tiene una distribución χ^2 con $n-1$ grados de libertad. Si deseamos probar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 > \sigma_0^2$, podemos usar $\chi^2 = (n-1)S^2/\sigma_0^2$ como nuestro estadístico de prueba, pero ¿cómo debemos seleccionar la región de rechazo RR?

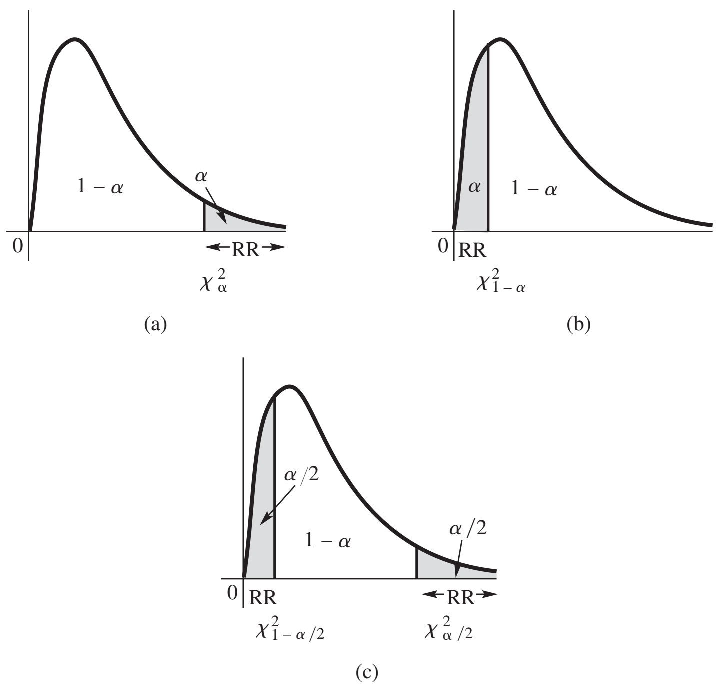
Si H_a es verdadera y el valor real de σ^2 es mayor que σ_0^2, esperaríamos que S^2 (que estima el verdadero valor de σ^2) fuera mayor que σ_0^2. Cuanto más grande sea S^2 con respecto a σ_0^2, más fuerte es la evidencia que apoye $H_a : \sigma^2 > \sigma_0^2$. Observe que S^2 es grande con respecto a σ_0^2 si y sólo si $\chi^2 = (n-1)S^2/\sigma_0^2$ es grande. Entonces, vemos que una región de rechazo de la forma RR $= \{\chi^2 > k\}$ para alguna constante k es apropiada para probar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 > \sigma_0^2$. Si deseamos una prueba para la cual la probabilidad de un error tipo I es α, usamos la región de rechazo

$$\text{RR} = \{\chi^2 > \chi_\alpha^2\},$$

donde $P(\chi^2 > \chi_\alpha^2) = \alpha$ (Valores de χ_α^2 se pueden hallar en la Tabla 6, Apéndice 3.) Una ilustración de esta región de rechazo se encuentra en la Figura 10.10(a).

Si deseamos probar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 < \sigma_0^2$ (una alternativa de cola inferior), un razonamiento análogo lleva a una región de rechazo ubicada en la cola inferior de la distribución χ^2. Por otra parte, podemos probar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 \neq \sigma_0^2$ (una prueba de dos colas) si usamos una región de rechazo de dos colas. Las gráficas que ilustran estas regiones de rechazo se dan en la Figura 10.10.

FIGURA **10.10** Regiones de rechazo RR para probar $H_0 : \sigma^2 = \sigma_0^2$ contra (a) $H_a : \sigma^2 > \sigma_0^2$; (b) $H_a : \sigma^2 < \sigma_0^2$; y (c) $H_a : \sigma^2 \neq \sigma_0^2$



Pruebas de hipótesis referentes a una varianza poblacional

Supuestos: $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con

$$E(Y_i) = \mu \text{ y } V(Y_i) = \sigma^2.$$

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2 & \text{(alternativa de cola superior).} \\ \sigma^2 < \sigma_0^2 & \text{(alternativa de cola inferior).} \\ \sigma^2 \neq \sigma_0^2 & \text{(alternativa de dos colas).} \end{cases}$$

Estadístico de prueba: $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$.

$$\text{Región de rechazo:} \quad \begin{cases} \chi^2 > \chi_\alpha^2 & \text{(RR de cola superior).} \\ \chi^2 < \chi_{1-\alpha}^2 & \text{(RR de cola inferior)} \\ \chi^2 > \chi_{\alpha/2}^2 \text{ o } \chi^2 < \chi_{1-\alpha/2}^2 & \text{(RR de dos colas).} \end{cases}$$

Observe que χ_α^2 se elige de modo que, para $\nu = n - 1$ grados de libertad, $P(\chi^2 > \chi_\alpha^2) = \alpha$ (Vea la Tabla 6, Apéndice 3.)

EJEMPLO **10.16** Una compañía produce piezas maquinadas para motor que se supone tienen una varianza en diámetro no mayor que .0002 (diámetros medidos en pulgadas). Una muestra aleatoria de diez piezas dio una varianza muestral de .0003. Pruebe, en el nivel de 5%, $H_0 : \sigma^2 = .0002$ contra $H_a : \sigma^2 > .0002$.

Solución Si es razonable suponer que los diámetros medidos están distribuidos normalmente, el estadístico de prueba adecuado es $\chi^2 = (n-1)S^2/\sigma_0^2$. Como hemos planteado una prueba de cola superior, rechazamos H_0 por valores de este estadístico mayores que $\chi_{.05}^2 = 16.919$ (con base en 9 grados de libertad). El valor observado del estadístico de prueba es

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{(9)(.0003)}{.0002} = 13.5.$$

Entonces, H_0 no es rechazada. No hay suficiente evidencia para indicar que σ^2 excede de .0002 en el nivel de significancia de 5%. ■

EJEMPLO **10.17** Determine el valor p relacionado con la prueba estadística del Ejemplo 10.16.

Solución El valor p es la probabilidad de que una variable aleatoria χ^2 con 9 grados de libertad sea mayor que el valor observado de 13.5. El área correspondiente a esta probabilidad está sombreada en la Figura 10.11. Al examinar la fila correspondiente a 9 grados de libertad en la

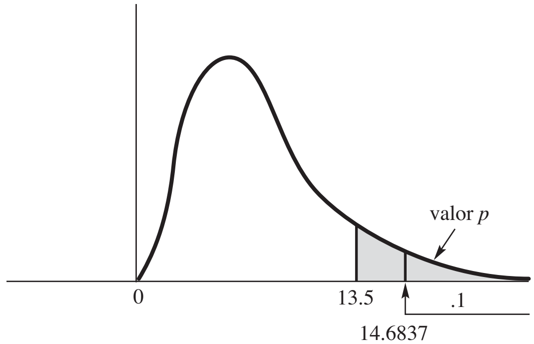


FIGURA **10.11** Ilustración del valor p para el Ejemplo 10.17 (densidad χ^2 con 9 grados de libertad)

Tabla 6, Apéndice 3, encontramos que $\chi^2_{.1} = 14.6837$. Como indica la Figura 10.11, el área sombreada excede de .1, y entonces el valor p es mayor que .1. Esto es, para cualquier valor de $\alpha < .1$, la hipótesis nula no puede ser rechazada. Esto está acorde con la conclusión del Ejemplo 10.16.

El valor exacto de p se obtiene fácilmente usando la aplicación *Chi-Square Probability and Quantiles*. Como se indica en la Figura 10.11, requerimos $P(\chi^2 > 13.5)$. Cuando χ^2 tiene 9 grados de libertad, como en la presente situación, la aplicación da $P(\chi^2 > 13.5) = .14126$. ■

EJEMPLO **10.18** Un experimentador está convencido de que la variabilidad en su equipo de medición resulta en una desviación estándar de 2. Dieciséis mediciones dieron como resultado $s^2 = 6.1$. ¿Los datos contradicen su afirmación? Determine el valor p para la prueba. ¿Qué concluiría usted si elige $\alpha = .05$?

Solución Requerimos una prueba de $H_0 : \sigma^2 = 4$ contra $H_a : \sigma^2 \neq 4$, una prueba de dos colas. El valor del estadístico de prueba es $\chi^2 = 15(6.1)/4 = 22.875$. Si consultamos la Tabla 6, Apéndice 3, vemos que, para 15 grados de libertad $\chi^2_{.05} = 24.9958$ y $\chi^2_{.10} = 22.3072$. Entonces, la parte del valor p que cae en la cola superior está entre .05 y .10. Como necesitamos tomar en cuenta un área correspondiente igual en la cola inferior (esta área también está entre .05 y .10), se deduce que $.1 <$ valor $p < .2$. Usando la aplicación *Chi-Square Probability and Quantiles* para calcular el valor exacto de p, obtenemos $P(\chi^2 > 22.8750) = .0868$, y ese valor $p = 2(.0868) = .1736$. Ya sea que usemos los límites obtenidos de la Tabla 6 o el valor p exacto obtenido de la aplicación, es evidente que el valor seleccionado de $\alpha = .05$ es menor que el valor p; por tanto, no podemos rechazar lo dicho por los experimentadores en el nivel de $\alpha = .05$. ■

A veces deseamos comparar las varianzas de dos distribuciones normales, particularmente al probar para determinar si son iguales. Estos problemas se encuentran al comparar la precisión de dos instrumentos de medición, la variación en las características de calidad de un producto manufacturado o la variación en las calificaciones para dos procedimientos de prueba. Por ejemplo, suponga que $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ y $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ son muestras aleatorias independientes de distribuciones normales con medias desconocidas y que $V(Y_{1i}) = \sigma_1^2$ y $V(Y_{2i}) = \sigma_2^2$, donde σ_1^2 y σ_2^2 son incógnitas. Suponga que deseamos probar la hipótesis nula $H_0 : \sigma_1^2 = \sigma_2^2$ contra la alternativa $H_a : \sigma_1^2 > \sigma_2^2$.

Como las varianzas muestrales S_1^2 y S_2^2 estiman las respectivas varianzas poblacionales, rechazamos H_0 a favor de H_a si S_1^2 es mucho mayor que S_1^2. Esto es, usamos una región de rechazo RR de la forma

$$\text{RR} = \left\{ \frac{S_1^2}{S_2^2} > k \right\},$$

donde k se elige de modo que la probabilidad de cometer un error tipo I sea α. El valor apropiado de k depende de la distribución de probabilidad del estadístico S_1^2/S_2^2. Observe que $(n_1-1)S_1^2/\sigma_1^2$ y $(n_2 - 1)S_2^2/\sigma_2^2$ son variables aleatorias independientes χ^2. De la Definición 7.3 se deduce que

$$F = \frac{(n_1 - 1)S_1^2}{\sigma_1^2(n_1 - 1)} \Bigg/ \frac{(n_2 - 1)S_2^2}{\sigma_2^2(n_2 - 1)} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

tiene una distribución F con $(n_1 - 1)$ grados de libertad en el numerador y $(n_2 - 1)$ grados de libertad en el denominador. Dada la hipótesis nula que $\sigma_1^2 = \sigma_2^2$, se deduce que $F = S_1^2/S_2^2$ y la región de rechazo RR dada antes es equivalente a RR $= \{F > k\} = \{F > F_\alpha\}$, donde $k = F_\alpha$ es el valor de la distribución F con $v_1 = (n_1 - 1)$ y $v_2 = (n_2 - 1)$ tal que $P(F > F_\alpha) = \alpha$. En la Tabla 7, Apéndice 3 se dan valores de F_α. Esta región de rechazo se muestra en la Figura 10.12.

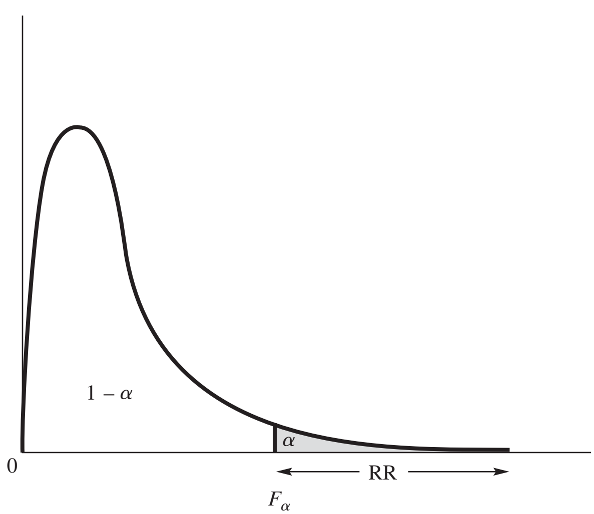


FIGURA **10.12** Región de rechazo RR para probar $H_0 : \sigma_1^2 = \sigma_2^2$ contra $H_a : \sigma_1^2 > \sigma_2^2$

EJEMPLO **10.19** Suponga que deseamos comparar la variación en los diámetros de las piezas producidas por la empresa del Ejemplo 10.16, con la variación en los diámetros de las piezas producidas por un competidor. Recuerde que la varianza muestral para nuestra compañía, basada en $n = 10$ diámetros, fue $s_1^2 = .0003$. En contraste, la varianza muestral de las mediciones de diámetro para 20 de las piezas del competidor fue $s_2^2 = .0001$. ¿Los datos proporcionan suficiente información para indicar una variación más pequeña en diámetros para el competidor? Pruebe usando $\alpha = .05$.

Solución Estamos probando $H_0 : \sigma_1^2 = \sigma_2^2$ contra la alternativa $H_a : \sigma_1^2 > \sigma_2^2$. El estadístico de prueba, $F = (S_1^2/S_2^2)$, está basado en $v_1 = 9$ grados de libertad en el numerador y $v_2 = 19$ en el denominador, y rechazamos H_0 para valores de F mayores que $F_{.05} = 2.42$. (Véase la Tabla 7, Apéndice 3.) Como el valor observado del estadístico de prueba es

$$F = \frac{s_1^2}{s_2^2} = \frac{.0003}{.0001} = 3,$$

vemos que $F > F_{.05}$; por tanto, en el nivel $\alpha = .05$, rechazamos $H_0 : \sigma_1^2 = \sigma_2^2$ a favor de $H_a : \sigma_1^2 > \sigma_2^2$ y concluimos que la compañía competidora produce piezas con menor variación en sus diámetros. ■

EJEMPLO **10.20** Limite el valor p relacionado con los datos del Ejemplo 10.19. Use la aplicación *F-Ratio Probabilities and Quantiles* para determinar el valor exacto de p.

Solución El valor F calculado para esta prueba de cola superior es $F = 3$. Como este valor está basado en $v_1 = 9$ y $v_2 = 19$ grados de libertad para numerador y denominador, respectivamente, la Tabla 7, Apéndice 3, se puede usar para determinar que $F_{.025} = 2.88$ mientras que $F_{.01} = 3.52$. Entonces, el valor observado, $F = 3$, llevaría al rechazo de la hipótesis nula para $\alpha = .025$ pero no para $\alpha = .01$. En consecuencia, $.01 <$ valor $p < .025$.

Requerimos el valor $p = P(F > 3)$ cuando F tiene una distribución F con $v_1 = 9$ grados de libertad para numerador y $v_2 = 19$ grados de libertad para denominador. El uso directo de la aplicación da como resultado $P(F > 3) = .02096$, un valor ubicado claramente entre .01 y .025, como está indicado por los límites para el valor p obtenido de la Tabla 7. ■

Suponga que, para el Ejemplo 10.19, nuestra hipótesis de investigación fue $H_a : \sigma_1^2 < \sigma_2^2$. ¿Cómo continuaríamos? Estamos en libertad de identificar cualquier población como la población 1. Por tanto, si intercambiamos simplemente las marcas arbitrarias de 1 y 2 en las dos poblaciones (y los correspondientes identificadores en tamaños muestrales, varianzas muestrales, etc.), nuestra hipótesis alternativa se convierte en $H_a : \sigma_1^2 > \sigma_2^2$, y podemos continuar como antes. Esto es, si la hipótesis de investigación es que la varianza de una población es mayor que la varianza en otra población, identificamos la población con la varianza *hipotética mayor* como población 1 y continuamos como se indica en la solución del Ejemplo 10.19.

Prueba de hipótesis $\sigma_1^2 = \sigma_2^2$

Suposiciones: muestras independientes de poblaciones normales.
$H_0: \sigma_1^2 = \sigma_2^2$.
$H_a: \sigma_1^2 > \sigma_2^2$.

Estadístico de prueba: $F = \dfrac{S_1^2}{S_2^2}$.

Región de rechazo: $F > F_\alpha$, donde F_α se elige para que $P(F > F_\alpha) = \alpha$ donde F tiene $v_1 = n_1 - 1$ grados de libertad en el numerador y $v_2 = n_2 - 1$ grados de libertad en el denominador. (Véase la Tabla 7, Apéndice 3.)

Si deseamos probar $H_0: \sigma_1^2 = \sigma_2^2$ contra $H_a: \sigma_1^2 \neq \sigma_2^2$ con la probabilidad α de cometer un error tipo I, podemos emplear $F = S_1^2/S_2^2$ como estadístico de prueba y rechazar H_0 a favor de H_a si el valor F calculado está en la cola superior o la inferior $\alpha/2$ de la distribución F. Los valores críticos de cola superior se pueden determinar directamente de la Tabla 7, Apéndice 3; pero, ¿cómo determinamos los valores críticos de cola inferior?

Observe que $F = S_1^2/S_2^2$ y $F^{-1} = S_2^2/S_1^2$ tienen distribuciones F, pero los grados de libertad en numerador y denominador se intercambian (el proceso de inversión cambia los papeles del numerador y el denominador). Denotemos con F_b^a una variable aleatoria con una distribución F con a y b grados de libertad en numerador y denominador, respectivamente, y sea $F_{b,\alpha/2}^a$ tal que

$$P\left(F_b^a > F_{b,\alpha/2}^a\right) = \alpha/2.$$

Entonces

$$P\left[\left(F_b^a\right)^{-1} < \left(F_{b,\alpha/2}^a\right)^{-1}\right] = \alpha/2$$

y, por tanto,

$$P\left[F_a^b < \left(F_{b,\alpha/2}^a\right)^{-1}\right] = \alpha/2.$$

Esto es, el valor que corta un área de cola inferior de $\alpha/2$ para una distribución F_a^b se puede hallar al invertir $F_{b,\alpha/2}^a$. Por tanto, si usamos $F = S_1^2/S_2^2$ como estadístico de prueba para probar $H_0: \sigma_1^2 = \sigma_2^2$ contra $H_a: \sigma_1^2 \neq \sigma_2^2$, la región de rechazo apropiada es

$$\text{RR}: \left\{F > F_{n_2-1,\alpha/2}^{n_1-1} \quad \text{o} \quad F < \left(F_{n_1-1,\alpha/2}^{n_2-1}\right)^{-1}\right\}.$$

Una prueba equivalente (vea el Ejercicio 10.81) se obtiene de la siguiente manera. Denote con n_L y n_S los tamaños muestrales relacionados con las varianzas muestrales mayor y menor, respectivamente. Ponga la varianza muestral mayor en el numerador y la varianza muestral menor en el denominador del estadístico F y rechace $H_0: \sigma_1^2 = \sigma_2^2$ a favor de $H_a: \sigma_1^2 \neq \sigma_2^2$ si $F > F_{\alpha/2}$, donde $F_{\alpha/2}$ está determinada por $v_1 = n_L - 1$ y $v_2 = n_S - 1$ grados de libertad en numerador y denominador, respectivamente.

EJEMPLO **10.21** Un experimento para explorar los umbrales del dolor provocados por descargas eléctricas para hombres y mujeres arrojó el resumen de datos que aparecen en la Tabla 10.4. ¿Los datos aportan suficiente evidencia para indicar una diferencia considerable en la variabilidad de umbrales del dolor para hombres y mujeres? Use $\alpha = .10$. ¿Qué se puede decir acerca del valor p?

Tabla 10.4 Datos para el Ejemplo 10.21

	Hombres	Mujeres
n	14	10
$\overline{y}$	16.2	14.9
s^2	12.7	26.4

Solución Supongamos que los umbrales del dolor para hombres y mujeres están distribuidos normalmente en forma aproximada. Deseamos probar $H_0 : \sigma_M^2 = \sigma_F^2$ contra $H_a : \sigma_M^2 \neq \sigma_F^2$, donde σ_M^2 y σ_F^2 son las varianzas de umbrales del dolor para hombres y mujeres, respectivamente. La S^2 mayor es 26.4 (la S^2 para mujeres), y el tamaño muestral relacionado con la S^2 mayor es $n_L = 10$. La S^2 menor es 12.7 (la S^2 para hombres) y $n_S = 14$ (el número de hombres de la muestra). Por tanto, calculamos

$$F = \frac{26.4}{12.7} = 2.079,$$

y comparamos este valor con $F_{\alpha/2} = F_{.05}$ con $v_1 = 10 - 1 = 9$ y $v_2 = 14 - 1 = 13$ grados de libertad en numerador y denominador, respectivamente. Dado que $F_{.05} = 2.71$ y como 2.079 no es mayor que el valor crítico (2.71), no existe evidencia suficiente para respaldar la afirmación de que la variabilidad de umbrales del dolor difiere para hombres y mujeres.

El valor p relacionado con el valor observado de F para esta prueba de dos colas se puede limitar de la siguiente manera. De acuerdo con la Tabla 7, Apéndice 3, con $v_1 = 9$, $v_2 = 13$ grados de libertad en numerador y denominador, respectivamente, encontramos $F_{.10} = 2.16$. Entonces, valor $p > 2(.10) = .20$. A menos que estemos dispuestos a trabajar con un valor muy grande de α (algún valor mayor que .2), estos resultados *no* nos permitirían concluir que las varianzas de los umbrales del dolor difieren para hombres y mujeres.

El valor p exacto se obtiene fácilmente usando la aplicación *F-Ratio Probabilities and Quantiles*. Con 9 grados de libertad en el numerador y 13 en el denominador, $P(F > 2.079) = .1005$ y el valor $p = 2(.1005) = .2010$, un valor mayor que .20, como se determina mediante el uso de la Tabla 7. ■

Aun cuando hemos empleado la notación F en el Ejemplo 10.21 para denotar la razón con la S^2 mayor en el numerador y la S^2 menor en el denominador, esta razón *no* tiene una distribución F (observe que la razón definida en esta forma *debe* ser mayor o igual a 1). Sin embargo, las tablas de la distribución F se pueden usar para determinar la región de rechazo para una prueba de nivel α (vea el Ejercicio 10.81).

Las pruebas de χ^2 y de F presentadas en esta sección son *muy sensibles* a desviaciones respecto a la suposición de normalidad de las población(es) subyacente(s). Entonces, a diferencia de las pruebas t de la Sección 10.8, estas pruebas *no son robustas* si se viola la suposición de normalidad.

Ejercicios

10.78 Un fabricante de cascos de seguridad para trabajadores de la construcción está interesado en la media y la varianza de las fuerzas que sus cascos transmiten a quienes los usan cuando se someten a una fuerza externa normal. El fabricante desea que la fuerza media transmitida por los cascos sea de 800 libras (o menos), bastante abajo del límite legal de 1000 libras, y desea que σ sea menor que 40. Se ejecutaron pruebas en una muestra aleatoria de $n = 40$ cascos, encontrando que la media muestral y la varianza eran iguales a 825 libras y 2350 libras^2, respectivamente.

a Si $\mu = 800$ y $\sigma = 40$, ¿es probable que cualquier casco sometido a la fuerza externa normal transmita una fuerza de más de 1000 libras al usuario? Explique.

b ¿Los datos proporcionan suficiente evidencia para indicar que cuando se someten a la fuerza externa normal, los cascos transmiten una fuerza media de más de 800 libras?

c ¿Los datos aportan suficiente evidencia para indicar que σ excede de 40?

10.79 Un fabricante de máquinas para empacar jabón en polvo afirma que su máquina podría cargar cajas con un peso dado y una variación de no más de .4 onzas. Se encontró que la media y la varianza de una muestra de ocho cajas de 3 libras fue de 3.1 y .018, respectivamente. Pruebe la hipótesis de que la varianza de la población de mediciones de peso es $\sigma^2 = .01$ contra la alternativa de que $\sigma^2 > .01$.

a Use un nivel de significancia de $\sigma = .05$. ¿Qué suposiciones se requieren para esta prueba?

b ¿Qué puede decirse acerca del nivel de significancia alcanzado usando una tabla del apéndice?

c **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia obtenido usando la aplicación apropiada?

10.80 ¿Con qué suposiciones puede usarse la distribución F para hacer inferencias acerca de la razón entre varianzas poblacionales?

10.81 En dos poblaciones normales con varianzas respectivas de σ_1^2 y σ_2^2, observamos varianzas muestrales independientes S_1^2 y S_2^2, con correspondientes grados de libertad $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$. Deseamos probar $H_0 : \sigma_1^2 = \sigma_2^2$ contra $H_a : \sigma_1^2 \neq \sigma_2^2$.

a Demuestre que la región de rechazo dada por

$$\left\{F > F_{\nu_2,\,\alpha/2}^{\nu_1} \quad \text{o} \quad F < \left(F_{\nu_1,\,\alpha/2}^{\nu_2}\right)^{-1}\right\},$$

donde $F = S_1^2/S_2^2$, es la misma región de rechazo dada por

$$\left\{S_1^2/S_2^2 > F_{\nu_2,\,\alpha/2}^{\nu_1} \quad \text{o} \quad S_2^2/S_1^2 > F_{\nu_1,\,\alpha/2}^{\nu_2}\right\}.$$

b Denote con S_L^2 la mayor de S_1^2 y S_2^2 y denote con S_S^2 la menor de S_1^2 y S_2^2. Denote con ν_L y ν_S los grados de libertad asociados con S_1^2 y S_2^2, respectivamente. Use el inciso a para demostrar que, de acuerdo con H_0,

$$P\left(S_L^2/S_S^2 > F_{\nu_S,\,\alpha/2}^{\nu_L}\right) = \alpha.$$

Observe que esto proporciona un método equivalente para probar la igualdad de dos varianzas.

10.82 Los Ejercicios 8.83 y 10.73 presentaron algunos datos recolectados en un estudio de 1993 por Susan Beckham y sus colegas. En este estudio, mediciones de la presión del compartimento anterior del músculo (en milímetros de mercurio) se tomaron para diez corredores sanos y diez ciclistas sanos. Los investigadores también obtuvieron mediciones de presión para los corredores y ciclistas al máximo consumo de O_2. El resumen de datos se da en la siguiente tabla.

Condición	Corredores		Ciclistas	
	Media	*s*	Media	*s*
Reposo	14.5	3.92	11.1	3.98
80% consumo máx. de O_2	12.2	3.49	11.5	4.95
Consumo máx. de O_2	19.1	16.9	12.2	4.67

a ¿Hay suficiente evidencia para apoyar la afirmación de que la variabilidad en presión del compartimento anterior del músculo difiere para corredores y ciclistas que estén en reposo? Use $\alpha = .05$.

b **i** ¿Qué se puede decir acerca del nivel de significancia obtenido usando una tabla del apéndice?

ii **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia obtenido usando la aplicación apropiada?

c ¿Hay suficiente evidencia para apoyar la afirmación de que la variabilidad en presión del compartimento anterior del músculo entre corredores y ciclistas difiere al máximo consumo de O_2? Use $\alpha =$ .05.

d **i** ¿Qué se puede decir acerca del nivel de significancia obtenido usando una tabla del apéndice?

ii **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia obtenido usando la aplicación apropiada?

10.83 El gerente de una lechería desea comprar una nueva máquina llenadora de botellas y está considerando las máquinas fabricadas por las compañías A y B. Si la robustez, el costo y la comodidad son similares en las dos máquinas, el factor decisivo será la variabilidad de llenados (es preferible la máquina que produzca llenados con menor varianza). Sean σ_1^2 y σ_2^2 las varianzas de llenado para máquinas producidas por las compañías A y B, respectivamente. Ahora considere varias pruebas de la hipótesis nula $H_0 : \sigma_1^2 = \sigma_2^2$. Al obtener muestras de llenados de las dos máquinas y usar el estadístico de prueba S_1^2/S_2^2, podríamos establecer como región de rechazo un área de cola superior, un área de cola inferior o un área de dos colas de la distribución *F*, dependiendo de los intereses a servir. Identifique el tipo de región de rechazo que sería más favorecido por las siguientes personas y explique por qué.

a El gerente de la lechería

b Un vendedor de la compañía A

c Un vendedor de la compañía B

10.84 Un experimento publicado en *The American Biology Teacher* estudió la eficacia de usar 95% de etanol y 20% de blanqueador como desinfectantes para eliminar contaminación por bacterias y hongos cuando se cultivan tejidos de plantas. El experimento se repitió 15 veces con cada uno de los desinfectantes, usando berenjenas como el tejido de planta cultivado.[17] Cinco cortes por planta se colocaron en una caja de petri desinfectada usando cada uno de los agentes y almacenada a 25°C durante 4 semanas. Las observaciones reportadas fueron el número de cortes de berenjena no contaminados después de 4 semanas de almacenamiento. La información relevante se da en la siguiente tabla. ¿Usted estaría dispuesto a suponer que las varianzas poblacionales subyacentes son iguales?

Desinfectante	95% etanol	20% blanqueador
Media	3.73	4.80
Varianza	2.78095	0.17143
n	15	15

17. *Fuente*: Michael Brehm, J. Buguliskis, D. Hawkins, E. Lee, D. Sabapathi, and R. Smith, "Determining Differences in Efficacy of Two Disinfectants Using *t* tests", *The American Biology Teacher* 58(2), (1996): 111.

a ¿Qué se puede decir acerca del nivel de significancia alcanzado usando la tabla F del apéndice?

b **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia obtenido usando la aplicación *F-Ratio Probabilities and Quantiles*?

c ¿Qué concluiría usted, con $\alpha = .02$?

10.85 **Ejercicio Applet** Un instrumento de precisión está garantizado para ser preciso con una variación de no más de 2 unidades. Una muestra de cuatro lecturas de un instrumento en un mismo objeto dio las medidas 353, 351, 351 y 355. Dé el nivel de significancia alcanzado para probar la hipótesis nula $\sigma = .7$ contra la hipótesis alternativa $\sigma > .7$.

10.86 Los exámenes de aptitud deben producir calificaciones con una gran cantidad de variación para que un administrador pueda distinguir entre personas con baja aptitud y otras de elevada aptitud. El examen estándar empleado por cierta industria ha producido calificaciones con desviación estándar de 10 puntos. Un nuevo examen se aplica a 20 posibles empleados y produce una desviación estándar muestral de 12 puntos. ¿Las calificaciones del nuevo examen son considerablemente más variables que las del examen estándar? Use $\alpha = .01$.

10.87 Consulte el Ejercicio 10.70. ¿Hay suficiente evidencia, en el nivel de significancia de 5%, para apoyar la conclusión de que la varianza en mediciones de niveles de DDT es mayor para los pelícanos jóvenes que para los polluelos?

10.10 Potencia de las pruebas y el lema de Neyman-Pearson

En las siguientes secciones de este capítulo, pasamos de ejemplos prácticos de pruebas estadísticas a una exposición teórica de sus propiedades. Hemos sugerido pruebas específicas para varias situaciones prácticas de pruebas de hipótesis, pero usted puede preguntarse por qué seleccionamos esas pruebas en particular. ¿Cómo decidimos sobre los estadísticos de prueba que se presentaron y cómo sabíamos que habíamos seleccionado las mejores regiones de rechazo?

La bondad de una prueba es medida por α y β, las probabilidades de errores tipo I y tipo II, respectivamente. Por lo general, el valor de α se elige de antemano y determina la ubicación de la región de rechazo. Un concepto relacionado pero muy útil para evaluar el desempeño de una prueba recibe el nombre de *potencia* de la prueba. Básicamente, la potencia de una prueba es la probabilidad de que la prueba lleve al rechazo de la hipótesis nula.

DEFINICIÓN **10.3**

Suponga que W es el estadístico de prueba y RR es la región de rechazo para una prueba de una hipótesis que involucra el valor de un parámetro θ. Entonces la *potencia* de la prueba, denotada por potencia(θ), es la probabilidad de que la prueba lleve al rechazo de H_0 cuando el valor real del parámetro es θ. Esto es,

$$\text{potencia}(\theta) = P\,(W \text{ en RR cuando el valor del parámetro es } \theta).$$

Suponga que deseamos probar la hipótesis nula $H_0 : \theta = \theta_0$ y que θ_a es un valor particular para θ elegido de H_a. La potencia de la prueba en $\theta = \theta_0$, potencia(θ_0), es igual a la probabilidad de rechazar H_0 cuando H_0 es verdadera. Esto es, potencia(θ_0) $= \alpha$, la probabilidad de cometer un error tipo I. Para cualquier valor de θ a partir de H_a, la potencia de una prueba mide

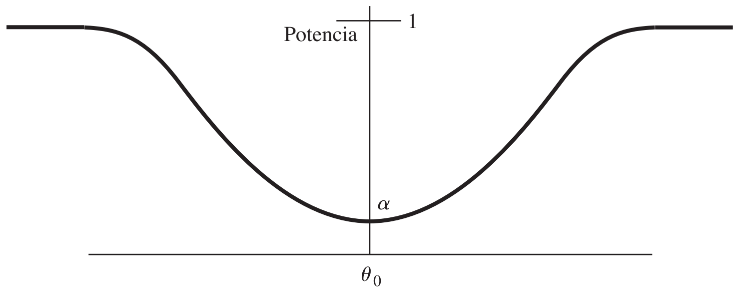


FIGURA **10.13** Curva típica de potencia para la prueba de $H_0 : \theta = \theta_0$ contra la alternativa $H_a : \theta \neq \theta_0$

la capacidad de la prueba para detectar que la hipótesis nula es falsa. Esto es, para $\theta = \theta_a$,

$$\text{potencia}(\theta_a) = P\,(\text{rechazar } H_0 \text{ cuando } \theta = \theta_a).$$

Si expresamos la probabilidad β de cometer un error tipo II cuando $\theta = \theta_a$ como $\beta(\theta_a)$, entonces

$$\beta(\theta_a) = P\,(\text{aceptar } H_0 \text{ cuando } \theta = \theta_a).$$

Se deduce que la potencia de la prueba en θ_a y la probabilidad de cometer un error tipo II están relacionados de la siguiente manera.

Relación entre potencia y β

Si θ_a es un valor de θ en la hipótesis alternativa H_a, entonces

$$\text{potencia}(\theta_a) = 1 - \beta(\theta_a).$$

Una *curva de potencia* típica, una gráfica de potencia(θ), se ilustra en la Figura 10.13.

Idealmente, una prueba detectaría una desviación desde $H_0 : \theta = \theta_0$ con certeza; esto es, potencia(θ_a) sería 1 para toda θ_a en H_a (vea la Figura 10.14). Porque, para un tamaño muestral fijo, α y β no pueden hacerse arbitrariamente pequeños, es evidente que esto no es posible. Por tanto, para un tamaño muestral fijo n, adoptamos el procedimiento de seleccionar un valor (pequeño) para α y hallar una región de rechazo RR para *minimizar* $\beta(\theta_a)$ en cada θ_a en H_a. De manera equivalente, seleccionamos RR para maximizar la potencia(θ) para θ en H_a. De entre todas las pruebas con nivel de significancia de α, buscamos la prueba cuya función de potencia se acerque más a la función de potencia ideal (Figura 10.14) si existe esa prueba. ¿Cómo determinamos este procedimiento de prueba?

Antes de continuar, debemos definir las hipótesis *simple* y *compuesta*. Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución exponencial con parámetro

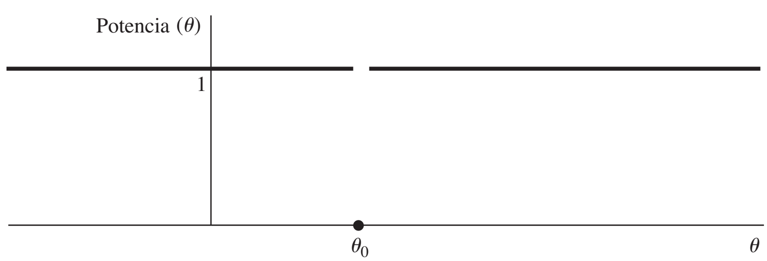


FIGURA **10.14** Curva ideal de potencia para la prueba de $H_0 : \theta = \theta_0$ contra $H_a : \theta \neq \theta_0$

λ; esto es, $f(y) = (1/y)e^{-y/\lambda}$, $y > 0$. Entonces la hipótesis $H : \lambda = 2$ especifica de manera única la distribución de la cual se toma la muestra y que tiene función de densidad $f(y) = (1/2)e^{-y/2}$, $y > 0$. La hipótesis $H : \lambda = 2$ es, por tanto, un ejemplo de una hipótesis *simple*. En contraste, la hipótesis $H^* : \lambda > 2$ es una hipótesis *compuesta* porque dada H^* la función de densidad $f(y)$ no está determinada de manera única. La forma de la densidad es exponencial, pero el parámetro λ podría ser 3 o 15 o cualquier valor mayor que 2.

DEFINICIÓN **10.4**

Si se toma una muestra aleatoria de una distribución con parámetro θ, se dice que una hipótesis es *simple* si *especifica de manera única* la distribución de la población de la cual se toma la muestra. Cualquier hipótesis que no sea simple se denomina *hipótesis compuesta*.

Si $Y_1, Y_2, \ldots, Y_n$ representan una muestra aleatoria de una distribución normal con varianza conocida $\sigma^2 = 1$, entonces $H : \mu = 5$ es una hipótesis simple porque, si H es verdadera, la función de densidad está especificada de manera única para ser una función de densidad normal con $\mu = 5$ y $\sigma^2 = 1$. Si, por otro lado, σ^2 no se conoce, la hipótesis $H : \mu = 5$ determina la media de la distribución normal pero no determina el valor de la varianza. Por tanto, si σ^2 no se conoce, $H : \mu = 5$ es una hipótesis compuesta.

Supongamos que nos gustaría probar una hipótesis nula *simple* $H_0 : \theta = \theta_0$ contra una hipótesis alternativa *simple* $H_a : \theta = \theta_a$. Como estamos interesados sólo en dos valores particulares de θ (θ_0 y θ_a), nos gustaría escoger una región de rechazo RR para que α = potencia(θ_0) sea un valor fijo y potencia(θ_a) sea tan grande como sea posible. Esto es, buscamos la *más potente* prueba de nivel α. El siguiente teorema proporciona la metodología para obtener la más potente prueba para probar H_0 simple contra H_a simple. [*Nota*: al igual que en la Definición 9.4, usamos la notación $L(\theta) = L(y_1, y_2, \ldots, y_n|\theta)$ para indicar que la función de verosimilitud depende de $y_1, y_2, \ldots, y_n$ y de θ.]

TEOREMA **10.1**

El lema de Neyman−Pearson Suponga que deseamos probar la hipótesis nula simple $H_0 : \theta = \theta_0$ contra la hipótesis alternativa simple $H_a : \theta = \theta_a$, con base en una muestra aleatoria $Y_1, Y_2, \ldots, Y_n$ de una distribución con parámetro θ. Sea $L(\theta)$ la verosimilitud de la muestra cuando el valor del parámetro es θ. Entonces, para una α dada, la prueba que maximiza la potencia en θ_a tiene una región de rechazo, RR, determinada por

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

El valor de k se escoge de modo que la prueba tenga el valor deseado para α. Esta prueba es la más potente en el nivel α para H_0 contra H_a.

La demostración del Teorema 10.1 no se incluye aquí, pero se puede encontrar en algunos de los textos citados en la bibliografía al final de este capítulo. Ilustramos la aplicación del teorema con el siguiente ejemplo.

EJEMPLO **10.22** Suponga que Y representa una sola observación de una población con función de densidad de probabilidad dada por

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Encuentre la más potente prueba con nivel de significancia $\alpha = .05$ para probar $H_0 : \theta = 2$ contra $H_a : \theta = 1$.

Solución Como ambas hipótesis son simples, el Teorema 10.1 puede aplicarse para obtener la prueba requerida. En este caso,

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{f(y|\theta_0)}{f(y|\theta_a)} = \frac{2y}{(1)y^0} = 2y, \qquad \text{para } 0 < \text{y} < 1,$$

y la forma de la región de rechazo para la más potente prueba es

$$2y < k.$$

Del mismo modo, la región de rechazo RR es $\{y < k/2\}$. O bien, debido a que $k/2 = k^*$ es una constante, la región de rechazo es RR: $\{y < k^*\}$.

Como $\alpha = .05$ está especificada, el valor de k^* está determinado por

$$.05 = P(Y \text{ en RR cuando } \theta = 2) = P(Y < k^* \text{ cuando } \theta = 2) = \int_0^{k^*} 2y \, dy = (k^*)^2.$$

Por tanto, $(k^*)^2 = .05$ y la región de rechazo de la más potente prueba es

$$\text{RR}: \{y < \sqrt{.05} = .2236\}$$

Entre todas las pruebas para H_0 contra H_a basadas en un tamaño muestral de 1 y con α fija en .05, esta prueba tiene el máximo valor posible para potencia(θ_a) = potencia(1). De la misma manera, entre todas las pruebas con $\alpha = .05$ esta prueba tiene la mínima probabilidad de cometer un error tipo II cuando $\beta(\theta_a)$ se evalúa en $\theta_a = 1$. ¿Cuál es el valor real para potencia(θ) cuando $\theta = 1$?

$$\begin{aligned} \text{potencia}(1) &= P(Y \text{ en RR cuando } \theta = 1) = P(Y < .2236 \text{ cuando } \theta = 1) \\ &= \int_0^{.2236} (1)\, dy = .2236 \end{aligned}$$

Aun cuando la región de rechazo $\{y < .2236\}$ da el *máximo* valor para potencia(1) entre todas las pruebas con $\alpha = .05$, vemos que $\beta(1) = 1 - .2236 = .7764$ todavía es muy grande. ■

Observe que las formas del estadístico de prueba y la región de rechazo dependen de H_0 y H_a. Si se cambia la alternativa a $H_a : \theta = 4$, la prueba más potente está basada en Y^2, y rechazamos H_0 a favor de H_a si $Y^2 > k'$, para alguna constante k'. También observe que el lema de Neyman−Pearson proporciona la *forma* de la región de rechazo; la región de rechazo real depende del valor especificado para α.

Para distribuciones discretas, no siempre es posible hallar una prueba cuyo nivel de significancia sea exactamente igual a algún valor predeterminado de α. En tales casos, especificamos que la prueba sea aquella para la cual la probabilidad de cometer un error tipo I sea la más cercana al valor predeterminado de α sin rebasarlo.

Suponga que hacemos un muestreo de una población cuya distribución está especificada por completo, excepto para el valor de un solo parámetro θ. Si deseamos probar $H_0: \theta = H_0$ (simple) contra $H_a: \theta > \theta_0$ (compuesta), ningún teorema comparable con el Teorema 10.1 es aplicable si cualquiera de las dos hipótesis es compuesta. No obstante, el Teorema 10.1 se puede aplicar para obtener una más potente prueba para $H_0 : \theta = \theta_0$ contra $H_a : \theta = \theta_a$ para cualquier valor individual θ_a, donde $\theta_a > \theta_0$. En muchas situaciones, la región de rechazo real para la más potente prueba depende sólo del valor de θ_0 (y no depende de la selección particular de θ_a). Cuando una prueba obtenida por el Teorema 10.1 en realidad maximiza la potencia para todo valor de θ mayor que θ_0, se dice que es una prueba *uniformemente más potente* para $H_0: \theta = \theta_0$ contra $H_a : \theta > \theta_0$. Observaciones análogas se aplican a la deducción de pruebas para $H_0: \theta = \theta_0$ contra $H_a : \theta < \theta_0$. Ilustramos estas ideas en el siguiente ejemplo.

EJEMPLO **10.23** Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 conocida. Deseamos probar $H_0: \mu = \mu_0$ contra $H_a: \mu > \mu_0$ para una constante especificada μ_0. Encuentre la prueba uniformemente más potente con un nivel de significancia α.

Solución Empezamos por buscar la prueba nivel α más potente de $H_0: \mu = \mu_0$ contra $H_a^*: \mu = \mu_a$ para un valor fijo de μ_a que es mayor que μ_0. Como

$$f(y \mid \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right], \qquad -\infty < y < \infty$$

tenemos

$$L(\mu) = f(y_1 \mid \mu)f(y_2 \mid \mu)\cdots f(y_n \mid \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{2\sigma^2}\right].$$

[Recuerde que exp(w) es simplemente e^w en otra forma.] Como H_0 y H_a^* son hipótesis *simples*, el Teorema 10.1 implica que la prueba más potente de $H_0: \mu = \mu_0$ contra $H_a^*: \mu = \mu_a$ está dada por

$$\frac{L(\mu_0)}{L(\mu_a)} < k,$$

que en este caso es equivalente a

$$\frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n}\frac{(y_i-\mu_0)^2}{2\sigma^2}\right]}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n}\frac{(y_i-\mu_a)^2}{2\sigma^2}\right]} < k.$$

Esta desigualdad se puede reacomodar como sigue:

$$\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2 - \sum_{i=1}^{n}(y_i-\mu_a)^2\right]\right\} < k.$$

Aplicando logaritmos naturales y simplificando, tenemos

$$-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu_0)^2-\sum_{i=1}^{n}(y_i-\mu_a)^2\right]<\ln(k)$$

$$\sum_{i=1}^{n}(y_i-\mu_0)^2-\sum_{i=1}^{n}(y_i-\mu_a)^2>-2\sigma^2\ln(k)$$

$$\sum_{i=1}^{n}y_i^2-2n\overline{y}\mu_0+n\mu_0^2-\sum_{i=1}^{n}y_i^2+2n\overline{y}\mu_a-n\mu_a^2>-2\sigma^2\ln(k)$$

$$\overline{y}(\mu_a-\mu_0)>\frac{-2\sigma^2\ln(k)-n\mu_0^2+n\mu_a^2}{2n}$$

o bien, como $\mu_a > \mu_0$,

$$\overline{y}>\frac{-2\sigma^2\ln(k)-n\mu_0^2+n\mu_a^2}{2n(\mu_a-\mu_0)}.$$

Como σ^2, n, μ_0 y μ_a son constantes conocidas todas ellas, la cantidad del lado derecho de la desigualdad es una constante; llamémosla k'. Por tanto, la prueba más potente de $H_0: \mu = \mu_0$ contra $H_a^*: \mu = \mu_a$ tiene la región de rechazo dada por

$$\text{RR} = \{\overline{y} > k'\}.$$

El valor preciso de k' se determina al fijar α y observar que

$$\begin{aligned}\alpha &= P(\overline{Y}\text{ en RR cuando }\mu=\mu_0)\\ &= P(\overline{Y}>k'\text{ cuando }\mu=\mu_0)\\ &= P\left(\frac{\overline{Y}-\mu_0}{\sigma/\sqrt{n}}>\frac{k'-\mu_0}{\sigma/\sqrt{n}}\right)\\ &= P\left(Z>\sqrt{n}(k'-\mu_0)/\sigma\right).\end{aligned}$$

Porque, dada H_0, Z tiene una distribución normal estándar, $P(Z > z_\alpha) = \alpha$ y el valor requerido para k' debe satisfacer

$$\sqrt{n}(k'-\mu_0)/\sigma = z_\alpha,\ \text{o bien, de manera equivalente,}\ k' = \mu_0 + z_\alpha\sigma/\sqrt{n}.$$

Entonces, la prueba nivel α que tiene el máximo valor posible para la potencia(θ_a) está basada en el estadístico $\overline{Y}$ y tiene región de rechazo $\text{RR} = \{\overline{y} > \mu_0 + z_\alpha\sigma/\sqrt{n}\}$. Ahora observamos que ni el estadístico de prueba ni la región de rechazo para esta prueba nivel α dependen del valor particular asignado a μ_a. Esto es, para cualquier valor de μ_a mayor que μ_0, obtenemos exactamente la misma región de rechazo. Por tanto, la prueba nivel α con la región de rechazo previamente dada tiene el máximo valor posible para potencia(μ_a) para *toda* $\mu_a > \mu_0$. Es la prueba *uniformemente más potente* para $H_0: \mu = \mu_0$ contra $H_a: \mu > \mu_0$. Ésta es exactamente la prueba que consideramos en la Sección 10.3. ■

De nuevo consideremos el caso en que la muestra aleatoria se toma de una distribución que está completamente especificada, excepto para el valor de un solo parámetro θ. Si deseamos obtener una prueba para $H_0: \theta \leq \theta_0$ contra $H_a: \theta > \theta_0$ (de manera que tanto H_0 como H_a sean hipótesis compuestas), ¿cómo procedemos? Suponga que usamos el método ilustrado en el Ejemplo 10.23 para hallar una prueba uniformemente más potente para $H_0' : \theta = \theta_0$ contra $H_a : \theta > \theta_0$. Si θ_1 es un valor fijo de θ que es menor que θ_0 y usamos la misma prueba para $H_0'' : \theta = \theta_1$ contra H_a, por lo general, α disminuirá y la potencia(θ_a) permanecerá sin cambio para toda θ_a en H_a. En otras palabras, si tenemos una buena prueba para discriminar entre H_0' y H_a, la misma prueba será todavía mejor para discriminar entre H_0'' y H_a. Para pruebas con hipótesis nulas compuestas de la forma $H_0 : \theta \le \theta_0$ (o $H_0 : \theta \ge \theta_0$), definimos el nivel α de significancia como la probabilidad de un error tipo I cuando $\theta = \theta_0$; esto es, $\alpha =$ potencia(θ_0). Por lo general, este valor para α es el máximo valor de la función de potencia para $\theta \le \theta_0$ (o $\theta \ge \theta_0$). Usando esta metodología, podemos demostrar que la prueba obtenida en el Ejemplo 10.23 para probar $H_0 : \theta = \theta_0$ contra $H_a : \theta > \theta_0$ también es la prueba uniformemente más potente nivel α para probar $H_0 : \theta \le \theta_0$ contra $H_a : \theta > \theta_0$.

En el Ejemplo 10.23 obtuvimos la prueba uniformemente más potente para $H_0 : \mu = \mu_0$ contra $H_a : \mu > \mu_0$ y encontramos que tiene la región de rechazo $\{\overline{y} > \mu_0 + z_\alpha \sigma/\sqrt{n})\}$. Si deseamos probar $H_0 : \mu = \mu_0$ contra $H_a : \mu < \mu_0$, cálculos análogos nos llevarían a $\{\overline{y} < \mu_0 - z_\alpha \sigma/\sqrt{n}\}$ como la región de rechazo para la prueba que es uniformemente más potente para toda $\mu_a < \mu_0$. Por tanto, si deseamos probar $H_0 : \mu = \mu_0$ contra $H_a : \mu \ne \mu_0$, ninguna región de rechazo individual proporciona la prueba más potente para todos los valores de $\mu_a \ne \mu_0$. Aun cuando hay algunas excepciones especiales, en casi todos los casos no existen pruebas de dos colas uniformemente más potentes. Por tanto, hay muchas hipótesis nulas y alternativas para las cuales no existen pruebas uniformemente más potentes.

El lema de Neyman−Pearson es inútil si deseamos probar una hipótesis acerca de un solo parámetro θ cuando la distribución muestreada contiene otros parámetros no especificados. Por ejemplo, podríamos probar $H_0 : \mu = \mu_0$ cuando la muestra se toma de una distribución normal con varianza σ^2 desconocida. En este caso, $H_0 : \mu = \mu_0$ no determina de manera única la forma de la distribución (porque σ^2 podría ser cualquier número no negativo) y por tanto no es una hipótesis simple. La siguiente sección presenta un método muy general y ampliamente usado para desarrollar pruebas de hipótesis. El método es particularmente útil cuando están presentes parámetros no especificados (llamados *parámetros de ruido*).

Ejercicios

10.88 Consulte el Ejercicio 10.2. Encuentre la potencia de la prueba para cada alternativa en a-d.

a $p = .4$.
b $p = .5$.
c $p = .6$.
d $p = .7$.
e Trace una gráfica de la función de potencia.

10.89 Consulte el Ejercicio 10.5. Encuentre la potencia de la prueba 1 para cada alternativa en a-e.

a $\theta = .1$.
b $\theta = .4$.
c $\theta = .7$.
d $\theta = 1$.
e Trace una gráfica de la función de potencia.

***10.90** Consulte el Ejercicio 10.5.

a Encuentre la potencia de la prueba 2 para cada una de las siguientes alternativas: $\theta = .1$, $\theta = .4$, $\theta = .7$ y $\theta = 1$.

b Trace una gráfica de la función de potencia.

c Compare la función de potencia del inciso b con la función de potencia que haya encontrado en el Ejercicio 10.89 (ésta es la función de potencia para la prueba 1, Ejercicio 10.5). ¿Qué se puede concluir acerca de la potencia de prueba 2 comparada con la potencia de prueba 1 *para toda* $\theta \geq 0$?

10.91 Sea $Y_1, Y_2, \ldots, Y_{20}$ una muestra aleatoria de tamaño $n = 20$ de una distribución normal con media μ desconocida y varianza $\sigma^2 = 5$ conocida. Deseamos probar $H_0: \mu = 7$ contra $H_a: \mu > 7$.

a Encuentre la prueba uniformemente más potente con nivel de significancia .05.

b Para la prueba del inciso a, encuentre la potencia en cada uno de los siguientes valores alternativos para $\mu : \mu_a = 7.5, 8.0, 8.5$ y 9.0.

c Trace una gráfica de la función de potencia.

10.92 Considere la situación descrita en el Ejercicio 10.91. ¿Cuál es el mínimo tamaño muestral para que una prueba de nivel $\alpha = .05$ tenga potencia de al menos .80 cuando $\mu = 8$?

10.93 Para una distribución normal con media μ y varianza $\sigma^2 = 25$, un experimentador desea probar $H_0: \mu = 10$ contra $H_a: \mu = 5$. Encuentre el tamaño muestral n para el cual la prueba más potente tendrá $\alpha = \beta = .025$.

10.94 Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen un muestra aleatoria de una distribución normal con media μ *conocida* y varianza σ^2 desconocida. Encuentre la prueba nivel α más potente de $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 = \sigma_1^2$, donde $\sigma_1^2 > \sigma_0^2$. Muestre que esta prueba es equivalente a una prueba χ^2. ¿La prueba es uniformemente más potente para $H_a : \sigma^2 > \sigma_0^2$?

10.95 Suponga que tenemos una muestra aleatoria de cuatro observaciones de la función de densidad

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{2\theta^3}\right) y^2 e^{-y/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

a Encuentre la región de rechazo para la prueba más potente de $H_0: \theta = \theta_0$ contra $H_a: \theta = \theta_a$, suponiendo que $\theta_a > \theta_0$. [*Sugerencia:* haga uso de la distribución χ^2.]

b ¿La prueba dada en el inciso a es uniformemente más potente para la alternativa $\theta > \theta_0$?

10.96 Suponga que Y es una muestra aleatoria de tamaño 1 desde una población con función de densidad

$$f(y \mid \theta) = \begin{cases} \theta y^{\theta - 1}, & 0 \leq y \leq 1, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

donde $\theta > 0$.

a Trace la función de potencia de la prueba con región de rechazo: $Y > .5$.

b Con base en la sola observación de Y, encuentre una prueba uniformemente más potente de tamaño α para probar $H_0: \theta = 1$ contra $H_a: \theta > 1$.

***10.97** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes y distribuidas idénticamente con función de probabilidad discreta dada por

	y		
	1	2	3
$p(y \mid \theta)$	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

donde $0 < \theta < 1$. Denote con N_i el número de observaciones iguales a i para $i = 1, 2, 3$.

a Obtenga la función de probabilidad $L(\theta)$ como función de N_1, N_2 y N_3.

b Encuentre la prueba más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta = \theta_a$, donde $\theta_a > \theta_0$. Demuestre que su prueba especifica que H_0 sea rechazada para ciertos valores de $2N_1 + N_2$.

c ¿Cómo determina el valor de k para que la prueba tenga nivel nominal α? No necesita hacer el cálculo real. Una descripción clara de cómo determinar k es adecuada.

d ¿La prueba obtenida en los incisos a-c es uniformemente más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta > \theta_0$? ¿Por qué sí o por qué no?

10.98 Sea $Y_1, \dots, Y_n$ una muestra aleatoria de la función de densidad de probabilidad dada por

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta}\right) m y^{m-1} e^{-y^m/\theta}, & y > 0, \\ 0, & \text{en cualquier otro punto,} \end{cases}$$

con m denotando una constante conocida.

a Encuentre la prueba uniformemente más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta > \theta_0$.

b Si la prueba del inciso a ha de tener $\theta_0 = 100$, $\alpha = .05$ y $\beta = .05$ cuando $\theta_a = 400$, encuentre el tamaño muestral apropiado y la región crítica.

10.99 Denote con $Y_1, Y_2, \dots, Y_n$ una muestra aleatoria de una población que tiene una distribución de Poisson con media λ.

a Encuentre la forma de la región de rechazo para una prueba más potente de $H_0: \lambda = \lambda_0$ contra $H_a: \lambda = \lambda_a$, donde $\lambda_a > \lambda_0$.

b Recuerde que $\sum_{i=1}^{n} Y_i$ tiene una distribución de Poisson con media $n\lambda$. Indique el modo en que esta información se puede utilizar para determinar constantes asociadas con la región de rechazo obtenida en el inciso a.

c ¿La prueba obtenida en el inciso a es uniformemente más potente para probar $H_0: \lambda = \lambda_0$ contra $H_a: \lambda > \lambda_0$? ¿Por qué?

d Encuentre la forma de la región de rechazo para una prueba más potente de $H_0: \lambda = \lambda_0$ contra $H_a: \lambda = \lambda_a$, donde $\lambda_a < \lambda_0$.

10.100 Sea $Y_1, Y_2, \dots, Y_n$ una muestra aleatoria de una población que tiene una distribución de Poisson con media λ_1. Denote con $X_1, X_2, \dots, X_m$ una muestra aleatoria independiente de una población que tiene una distribución de Poisson con media λ_2. Obtenga la prueba más potente para probar $H_0: \lambda_1 = \lambda_2 = 2$ contra $H_a: \lambda_1 = 1/2, \lambda_2 = 3$.

10.101 Suponga que $Y_1, Y_2, \dots, Y_n$ denotan una muestra aleatoria de una población que tiene una distribución exponencial con media θ.

a Obtenga la prueba más potente para $H_0: \theta = \theta_0$ contra $H_a: \theta = \theta_a$, donde $\theta_a < \theta_0$.

b ¿La prueba obtenida en el inciso a es uniformemente más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta < \theta_0$?

10.102 Sea $Y_1, Y_2, \dots, Y_n$ una muestra aleatoria de una población con distribución de Bernoulli y parámetro p. Esto es,

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1.$$

a Suponga que estamos interesados en probar $H_0: p = p_0$ contra $H_a: p = p_a$, donde $p_0 < p_a$.

i Demuestre que

$$\frac{L(p_0)}{L(p_a)} = \left[\frac{p_0(1-p_a)}{(1-p_0)p_a}\right]^{\sum y_i} \left(\frac{1-p_0}{1-p_a}\right)^n.$$

ii Pruebe que $L(p_0)/L(p_a) < k$ si y sólo si $\sum_{i=1}^{n} y_i > k^*$ para alguna constante k^*.

iii Defina la región de rechazo para la prueba más potente de H_0 contra H_a.

b Recuerde que $\sum_{i=1}^{n} Y_i$ tiene una distribución binomial con parámetros n y p. Indique el modo de determinar los valores de cualesquiera constantes contenidas en la región de rechazo obtenidas en el inciso [a(iii)].

c ¿La prueba obtenida en el inciso a es uniformemente más potente para probar $H_0: p = p_0$ contra $H_a: p > p_0$? ¿Por qué sí o por qué no?

***10.103** Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución uniforme en el intervalo $(0, \theta)$.

a Encuentre la prueba nivel α más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta = \theta_a$, donde $\theta_a < \theta_0$.

b ¿La prueba obtenida en el inciso a es uniformemente más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta < \theta_0$?

***10.104** Consulte la muestra aleatoria del Ejercicio 10.103.

a Encuentre la prueba nivel α más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta = \theta_a$, donde $\theta_a > \theta_0$.

b ¿La prueba obtenida en el inciso a es uniformemente más potente para probar $H_0: \theta = \theta_0$ contra $H_a: \theta > \theta_0$?

c ¿Es única la prueba nivel α más potente que encontró en el inciso a?

10.11 Pruebas de razón de probabilidad

El Teorema 10.1 proporciona un método para construir pruebas más potentes para hipótesis simples cuando se conoce la distribución de las observaciones, excepto para el valor de un parámetro individual desconocido. Este método puede usarse a veces para hallar pruebas uniformemente más potentes para hipótesis compuestas que comprenden un parámetro individual. En muchos casos, la distribución de interés tiene más de un parámetro desconocido. En esta sección presentamos un método muy general que se puede usar para obtener pruebas de hipótesis. El procedimiento funciona para hipótesis simples o compuestas ya sea que estén presentes o no otros parámetros con valores desconocidos.

Suponga que una muestra aleatoria se selecciona de una distribución y que la función de probabilidad $L(y_1, y_2, \ldots, | y_n \mid \theta_1, \theta_2, \ldots, \theta_k)$ es una función de k parámetros, $\theta_1, \theta_2, \ldots, \theta_k$. Para simplificar la notación, denotemos con Θ el vector de todos los parámetros k, es decir, $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$, y escribimos la función de probabilidad como $L(\Theta)$. Puede ser el caso que estemos interesados en probar hipótesis sólo alrededor de uno de los parámetros, por ejemplo θ_1. Por ejemplo, si, como en el Ejemplo 10.24, tomamos una muestra de una población distribuida normalmente con media desconocida μ y varianza σ^2 desconocida, entonces la función de probabilidad depende de los *dos* parámetros μ y σ^2 y $\Theta = (\mu, \sigma^2)$. Si estamos interesados en probar hipótesis acerca de sólo la media μ, entonces σ^2, que es un parámetro no de particular interés para nosotros, se denomina *parámetro de ruido*. Por tanto, la función de probabilidad puede ser una función con parámetros de ruido desconocidos y un parámetro de interés.

Supongamos que la hipótesis nula especifica que Θ (puede ser un vector) se encuentra en un conjunto particular de posibles valores, por ejemplo Ω_0, y que la hipótesis alternativa especifica que Θ está en otro conjunto de posibles valores Ω_a, que no se traslapa con Ω_0. Por ejemplo, si muestreamos de una población con una distribución exponencial con media λ (en este caso, λ es el único parámetro de la distribución y $\Theta = \lambda$), podríamos estar interesados en probar H_0:

$\lambda = \lambda_0$ contra $H_a: \lambda \neq \lambda_0$. En este ejemplo exponencial, Ω_0 contiene sólo el valor individual λ_0 y $\Omega_a = \{\lambda > 0 : \lambda \neq \lambda_0\}$. Denote la unión de los dos conjuntos, Ω_0 y Ω_a, por Ω; esto es, $\Omega = \Omega_0 \cup \Omega_a$. En este ejemplo exponencial, $\Omega = \{\lambda_0\} \cup \{\lambda > 0 : \lambda \neq \lambda_0\} = \{\lambda : \lambda > 0\}$, es el conjunto de todos los posibles valores para λ. Cualquiera de las dos hipótesis H_0 y H_a, o ambas, pueden ser compuestas porque podrían contener valores múltiples del parámetro de interés o porque puedan estar presentes otros parámetros desconocidos.

Denotemos con $L(\hat{\Omega}_0)$ el máximo (en realidad el supremo) de la función de verosimilitud para toda $\Theta \in \Omega_0$. Esto es, $L(\hat{\Omega}_0) = \text{máx}_{\Theta \in \Omega_0} L(\Theta)$. Observe que $L(\hat{\Omega}_0)$ representa la mejor explicación para los datos observados para toda $\Theta \in \Omega_0$ y puede hallarse mediante métodos similares a los empleados en la Sección 9.7. Del mismo modo, $L(\hat{\Omega}) = \text{máx}_{\Theta \in \Omega} L(\Theta)$ representa la mejor explicación para los datos observados para toda $\Theta \in \Omega = \Omega_0 \cup \Omega_a$. Si $L(\hat{\Omega}_0) = L(\hat{\Omega})$, entonces una mejor explicación para los datos observados se puede hallar dentro de Ω_0, y no deberíamos rechazar la hipótesis nula $H_0: \Theta \in \Omega_0$; pero si $L(\hat{\Omega}_0) < L(\hat{\Omega})$, entonces la mejor explicación para los datos observados se puede hallar dentro de Ω_a, y deberíamos considerar rechazar H_0 a favor de H_a. Una prueba de razón de probabilidad está basada en la razón $L(\hat{\Omega}_0)/L(\hat{\Omega})$.

Prueba de razón de probabilidades

Defina λ mediante

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{\underset{\Theta \in \Omega_0}{\text{máx}}\, L(\Theta)}{\underset{\Theta \in \Omega}{\text{máx}}\, L(\Theta)}.$$

La prueba de razón de verosimilitudes de $H_0: \Theta \in \Omega_0$ contra $H_a: \Theta \in \Omega_a$ emplea λ como un estadístico de prueba, y la región de rechazo está determinada por $\lambda \leq k$.

Se puede demostrar que $0 \leq \lambda \leq 1$. Un valor de λ cercano a cero indica que la probabilidad de la muestra es mucho menor con H_0 que con H_a. Por tanto, los datos sugieren favorecer H_a sobre H_0. El valor real de k se selecciona de modo que α alcance el valor deseado. Ilustramos la mecánica de este método con el siguiente ejemplo.

EJEMPLO **10.24** Suponga que $Y_1, Y_2, \ldots, Y_n$ constituyen una muestra aleatoria de una distribución normal con media μ desconocida y varianza σ^2 desconocida. Deseamos probar $H_0: \mu = \mu_0$ contra $H_a: \mu > \mu_0$. Encuentre la prueba de razón de probabilidades adecuada.

Solución En este caso, $\Theta = (\mu, \sigma^2)$. Observe que Ω_0 es el conjunto $\{(\mu_0, \sigma^2) : \sigma^2 > 0\}$, $\Omega_a = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$, y por tanto que $\Omega = \Omega_0 \cup \Omega_a = \{(\mu, \sigma^2) : \mu \geq \mu_0,\ \sigma^2 > 0\}$. El valor constante de la varianza σ^2 es completamente no especificado. Debemos ahora hallar $L(\hat{\Omega}_0)$ y $L(\hat{\Omega})$.

Para la distribución normal tenemos

$$L(\Theta) = L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}\right].$$

Restringir μ a Ω_0 implica que $\mu = \mu_0$, y podemos hallar $L(\hat{\Omega}_0)$ si determinamos el valor de σ^2 que maximice $L(\mu, \sigma^2)$ sujeto a la restricción de que $\mu = \mu_0$. Del Ejemplo 9.15, vemos que cuando $\mu = \mu_0$ el valor de σ^2 que maximiza $L(\mu_0, \sigma^2)$ es

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2.$$

Entonces, $L(\hat{\Omega}_0)$ se obtiene al sustituir μ con μ_0 y σ^2 con $\hat{\sigma}_0^2$ en $L(\mu, \sigma^2)$, que da

$$L(\hat{\Omega}_0) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n}\frac{(y_i - \mu_0)^2}{2\hat{\sigma}_0^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} e^{-n/2}.$$

Ahora nos concentramos en hallar $L(\hat{\Omega})$. Al igual que en el Ejemplo 9.15, es más fácil ver en $\ln L(\mu, \sigma^2)$,

$$\ln[L(\mu, \sigma^2)] = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Evaluando derivadas con respecto a μ y σ^2, obtenemos

$$\begin{aligned}
\frac{\partial\{\ln[L(\mu, \sigma^2)]\}}{\partial\mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu), \\
\frac{\partial\{\ln[L(\mu, \sigma^2)]\}}{\partial\sigma^2} &= -\left(\frac{n}{2\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2.
\end{aligned}$$

Necesitamos hallar el máximo de $L(\mu, \sigma^2)$ en el conjunto $\Omega = \{(\mu, \sigma^2) : \mu \geq \mu_0, \sigma^2 > 0\}$. Observe que

$$\begin{aligned}
&\partial L(\mu, \sigma^2)/\partial\mu < 0, \qquad \text{si } \mu > \overline{y}, \\
&\partial L(\mu, \sigma^2)/\partial\mu = 0, \qquad \text{si } \mu = \overline{y}, \\
&\partial L(\mu, \sigma^2)/\partial\mu > 0, \qquad \text{si } \mu < \overline{y}.
\end{aligned}$$

Entonces, en el conjunto $\Omega = \{(\mu, \sigma^2) : \mu \geq \mu_0, \sigma^2 > 0\}$, $\ln L(\mu, \sigma^2)$ [y también $L(\mu, \sigma^2)$] se maximiza en $\hat{\mu}$ donde

$$\hat{\mu} = \begin{cases} \overline{y}, & \text{si } \overline{y} > \mu_0, \\ \mu_0, & \text{si } \overline{y} \leq \mu_0. \end{cases}$$

Al igual que antes, el valor de σ^2 en Ω que maximiza $L(\mu, \sigma^2)$, es

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2.$$

$L(\hat{\Omega})$ se obtiene al sustituir μ con $\hat{\mu}$ y σ^2 con $\hat{\sigma}^2$, que da

$$L(\hat{\Omega}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} \exp\left[-\sum_{i=1}^{n}\frac{(y_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right] = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}^2}\right)^{n/2} e^{-n/2}.$$

Entonces,

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$$

$$= \begin{cases} \left[\dfrac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2}\right]^{n/2}, & \text{si } \overline{y} > \mu_0 \\ 1, & \text{si } \overline{y} \leq \mu_0. \end{cases}$$

Observe que λ es siempre menor que o igual a 1. Así, valores "pequeños" de λ son aquellos que son menores que $k < 1$. Porque

$$\sum_{i=1}^n (y_i - \mu_0)^2 = \sum_{i=1}^n [(y_i - \overline{y}) + (\overline{y} - \mu_0)]^2$$
$$= \sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2$$

si $k < 1$, se deduce que la región de rechazo, $\lambda \leq k$, es equivalente a

$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2} < k^{2/n} = k'$$

$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2} < k'$$

$$\frac{1}{1 + \dfrac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2}} < k'.$$

Esta desigualdad, a su vez, es equivalente a

$$\frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2} > \frac{1}{k'} - 1 = k''$$

$$\frac{n(\overline{y} - \mu_0)^2}{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (y_i - \overline{y})^2} > (n-1)k''$$

o bien, como $\overline{y} > \mu_0$ cuando $\lambda < k < 1$,

$$\frac{\sqrt{n}(\overline{y} - \mu_0)}{s} > \sqrt{(n-1)k''},$$

donde

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2.$$

Observe que $\sqrt{n}(\overline{Y} - \mu_0)/S$ es el estadístico t empleado en secciones anteriores. En consecuencia, la prueba de razón de probabilidades es equivalente a la prueba t de la Sección 10.8. ■

No son raras las situaciones en las que la prueba de razón de probabilidades toma una forma bien conocida. De hecho, todas las pruebas de las Secciones 10.8 y 10.9 se pueden obtener con el método de razón de probabilidades. Para casi todos los problemas prácticos, el método de razón de probabilidades produce la mejor prueba posible en términos de potencia.

Desafortunadamente, el método de razón de probabilidades no siempre produce un estadístico de prueba con distribución de probabilidad conocida, por ejemplo el estadístico t del Ejemplo 10.24. Sin embargo, si el tamaño muestral es grande, podemos obtener una aproximación a la distribución de λ si algunas "condiciones de regularidad" razonables son satisfechas por la(s) distribución(es) poblacional(es) básica(s). Éstas son condiciones generales que se cumplen para la mayor parte (pero no todas) las distribuciones que hemos considerado. Las condiciones de regularidad comprenden básicamente la existencia de derivadas, con respecto a los parámetros, de la función de probabilidad. Otra condición clave es que la región sobre la cual la función de probabilidad es positiva no puede depender de valores paramétricos desconocidos.

TEOREMA 10.2 Sean $Y_1, Y_2, \dots, Y_n$ que tienen una función de probabilidad conjunta $L(\Theta)$. Denotemos con r_0 el número de parámetros libres que están especificados por $H_0: \Theta \in \Omega_0$ y denotemos con r el número de parámetros libres especificados por el enunciado $\Theta \in \Omega$. Entonces, para n grande, $-2\ln(\lambda)$ tiene aproximadamente una distribución χ^2 con $r_0 - r$ grados de libertad.

La prueba de este resultado está fuera del propósito de este libro. El Teorema 10.2 nos permite usar la tabla de la distribución χ^2 para determinar regiones de rechazo con α fija cuando n es grande. Observe que $-2\ln(\lambda)$ es una función decreciente de λ. Como la prueba de razón de probabilidades especifica que usemos RR: $\{\lambda < k\}$, este rechazo se puede reescribir como RR:$\{-2\ln(\lambda) > -2\ln(k) = k^*\}$. Para tamaños muestrales grandes, si deseamos una prueba de nivel α, el Teorema 10.2 implica que $k^* \approx \chi^2_\alpha$. Esto es, una prueba de razón de probabilidades para una muestra grande tiene la región de rechazo dada por

$$-2\ln(\lambda) > \chi^2_\alpha, \qquad \text{donde } \chi^2_\alpha \text{ está basada en } r_0 - r \text{ grados de libertad.}$$

El tamaño de la muestra necesaria para una "buena" aproximación varía de una aplicación a otra. Es importante darse cuenta que pruebas de razón de probabilidades para una muestra grande están basadas en $-2\ln(\lambda)$, donde λ es la razón de probabilidades *original*, $\lambda = L(\hat{\Omega}_0)/L(\hat{\Omega})$.

EJEMPLO 10.25 Suponga que un ingeniero desea comparar el número de quejas por semana que los representantes del sindicato de dos turnos diferentes en una planta manufacturera registran por semana. Cien observaciones independientes del número de quejas arrojó medias de $\overline{x} = 20$ por el turno 1 y $\overline{y} = 22$ por el turno 2. Suponga que el número de quejas por semana en el i-ésimo turno tiene una distribución de Poisson con media θ_i, para $i = 1, 2$. Use el método de razón de probabilidades para probar $H_0: \theta_1 = \theta_2$ contra $H_a: \theta_1 \neq \theta_2$ con $\alpha \approx .01$.

Solución La probabilidad de la muestra es ahora la función de probabilidad conjunta de todas las x_i y las y_j y está dada por

$$L(\theta_1, \theta_2) = \left(\frac{1}{k}\right)\theta_1^{\sum x_i} e^{-n\theta_1}\theta_2^{\sum y_j} e^{-n\theta_2},$$

donde $k = x_1! \cdots x_n!\, y_1! \cdots y_n!$ y $n = 100$. En este ejemplo, $\Theta = (\theta_1, \theta_2)$ y $\Omega_0 = \{(\theta_1, \theta_2): \theta_1 = \theta_2 = \theta\}$, donde θ es desconocida. Por tanto, dada H_0 la función de probabilidad es una función del parámetro individual θ, y

$$L(\theta) = \left(\frac{1}{k}\right)\theta^{\sum x_i + \sum y_j} e^{-2n\theta}.$$

Observe que, para $\Theta \in \Omega_0$, $L(\theta)$ se maximiza cuando θ es igual a su estimación de verosimilitud máxima,

$$\hat{\theta} = \frac{1}{2n}\left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{n} y_j\right) = \frac{1}{2}(\overline{x} + \overline{y}).$$

En este ejemplo $\Omega_a = \{(\theta_1, \theta_2) : \theta_1 \neq \theta_2\}$ y $\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$. Usando la probabilidad general $L(\theta_1, \theta_2)$, una función de θ_1 y de θ_2, vemos que $L(\theta_1, \theta_2)$ se maximiza cuando $\hat{\theta}_1 = \overline{x}$ y $\hat{\theta}_2 = \overline{y}$, respectivamente. Esto es, $L(\theta_1, \theta_2)$ se maximiza cuando θ_1 y θ_2 son sustituidas por sus estimaciones de máxima probabilidad. Entonces,

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{k^{-1}(\hat{\theta})^{n\overline{x}+n\overline{y}} e^{-2n\hat{\theta}}}{k^{-1}(\hat{\theta}_1)^{n\overline{x}}(\hat{\theta}_2)^{n\overline{y}} e^{-n\hat{\theta}_1 - n\hat{\theta}_2}} = \frac{(\hat{\theta})^{n\overline{x}+n\overline{y}}}{(\overline{x})^{n\overline{x}}(\overline{y})^{n\overline{y}}}.$$

Observe que λ es una función complicada de $\overline{x}$ y $\overline{y}$. El valor observado de $\hat{\theta}$ es $(1/2)(\overline{x} + \overline{y}) = (1/2)(20 + 22) = 21$. El valor observado de λ es

$$\lambda = \frac{21^{(100)(20+22)}}{20^{(100)(20)}22^{(100)(22)}}$$

y por tanto

$$-2\ln(\lambda) = -(2)[4200\ln(21) - 2000\ln(20) - 2200\ln(22)] = 9.53.$$

En esta aplicación, el número de parámetros libres en $\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}$ es $k = 2$. En $\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$, $r_0 = 1$ de estos parámetros libres es fijo. En el conjunto Ω, $r = 0$ de los parámetros son fijos. El Teorema 10.2 implica que $-2\ln(\lambda)$ tiene una distribución χ^2 aproximadamente con $r_0 - r = 1 - 0 = 1$ grado de libertad. Valores pequeños de λ corresponden a grandes valores de $-2\ln(\lambda)$, de modo que la región de rechazo para una prueba en aproximadamente el nivel $\alpha = .01$ contiene los valores de $-2\ln(\lambda)$ que excedan de $\chi^2_{.01} = 6.635$, el valor que corta un área de .01 en la cola derecha de una densidad χ^2 con 1 grado de libertad.

Como el valor observado de $-2\ln(\lambda)$ es mayor que $\chi^2_{.01}$, rechazamos $H_0 : \theta_1 = \theta_2$. Concluimos, en aproximadamente el nivel de significancia $\alpha = .01$, que los números medios de quejas presentadas por los representantes del sindicato difieren. ■

Ejercicios

10.105 Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución normal con media μ (desconocida) y varianza σ^2. Para probar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 > \sigma_0^2$, demuestre que la prueba de razón de probabilidades es equivalente a la prueba χ^2 dada en la Sección 10.9.

10.106 Una encuesta de la opinión de votantes fue realizada en cuatro delegaciones políticas del centro de una ciudad para comparar la fracción de votantes que están a favor del candidato A. Muestras aleatorias de 200 votantes se sondearon en cada una de las cuatro delegaciones, con los resultados que se muestran en la siguiente tabla. Los números de votantes que están a favor de A en las cuatro muestras se pueden registrar como cuatro variables aleatorias binomiales independientes.

Construya una prueba de razón de probabilidades de la hipótesis de que las fracciones de votantes que están a favor del candidato A son iguales en las cuatro delegaciones. Use $\alpha = .05$.

	Delegación				
Opinión	1	2	3	4	Total
A favor de A	76	53	59	48	236
No a favor de A	124	147	141	152	564
Total	200	200	200	200	800

10.107 Sean S_1^2 y S_2^2, respectivamente, las varianzas de muestras aleatorias independientes de tamaños n y m seleccionadas de distribuciones normales con medias μ_1 y μ_2 y varianza común σ^2. Si μ_1 y μ_2 son desconocidas, construya una prueba de razón de probabilidad de $H_0 : \sigma^2 = \sigma_0^2$ contra $H_a : \sigma^2 = \sigma_a^2$, suponiendo que $\sigma_a^2 > \sigma_0^2$.

10.108 Suponga que $X_1, X_2, \ldots, X_{n_1}$, $Y_1, Y_2, \ldots, Y_{n_2}$ y $W_1, W_2, \ldots, W_{n_3}$ son muestras aleatorias independientes de distribuciones normales con sus respectivas medias desconocidas μ_1, μ_2 y μ_3 y varianzas $\sigma_1^2, \sigma_2^2,$ y σ_3^2.

a Encuentre la prueba de razón de probabilidad para $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ contra la alternativa de al menos una desigualdad.

b Encuentre una región crítica aproximada para la prueba del inciso a si n_1, n_2 y n_3 son grandes y $\alpha = .05$.

***10.109** Sea $X_1, X_2, \ldots, X_m$ una muestra aleatoria de la densidad exponencial con media θ_1 y sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria independiente de una densidad exponencial con media θ_2.

a Encuentre el criterio de razón de probabilidad para probar $H_0 : \theta_1 = \theta_2$ contra $H_a : \theta_1 \neq \theta_2$.

b Demuestre que la prueba del inciso a es equivalente a una prueba F exacta [*Sugerencia:* transforme $\sum X_i$ y $\sum Y_j$ en variables aleatorias χ^2.]

***10.110** Demuestre que la prueba de razón de probabilidades depende de los datos sólo a través del valor de un estadístico suficiente. [*Sugerencia*: use el criterio de factorización.]

10.111 Suponga que estamos interesados en probar la hipótesis nula *simple* $H_0 : \theta = \theta_0$ contra la hipótesis alternativa *simple* $H_a : \theta = \theta_a$. De acuerdo con el lema de Neyman−Pearson, la prueba que maximiza la potencia en θ_a tiene una región de rechazo determinada por

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

En el contexto de una prueba de razón de probabilidades, si estamos interesados en las hipótesis *simples* H_0 y H_a, como se expresa, entonces, $\Omega_0 = \{\theta_0\}$, $\Omega_a = \{\theta_a\}$ y $\Omega = \{\theta_0, \theta_a\}$.

a Demuestre que la razón de probabilidades λ está dada por

$$\lambda = \frac{L(\theta_0)}{\text{máx}\{L(\theta_0),\, L(\theta_a)\}} = \frac{1}{\text{máx}\left\{1,\, \dfrac{L(\theta_a)}{L(\theta_0)}\right\}}.$$

b Demuestre que $\lambda < k$ si y sólo si, para alguna constante k',

$$\frac{L(\theta_0)}{L(\theta_a)} < k'.$$

c ¿Qué implican los resultados de los incisos a y b acerca de las pruebas de razón de probabilidades cuando la hipótesis nula y la alternativa son simples?

10.112 Suponga que muestras aleatorias independientes de tamaños n_1 y n_2 se han de seleccionar de poblaciones normales con medias μ_1 y μ_2, respectivamente, y varianza común σ^2. Para probar $H_0: \mu_1 = \mu_2$ contra $H_a: \mu_1 - \mu_2 > 0$ (σ^2 desconocida), demuestre que la prueba de razón de probabilidades se reduce a la prueba t de dos muestras presentada en la Sección 10.8.

10.113 Consulte el Ejercicio 10.112. Demuestre que para la prueba de $H_0: \mu_1 = \mu_2$ contra $H_a: \mu_1 \neq \mu_2$ (σ^2 desconocida) la prueba de razón de probabilidades se reduce a la prueba t de dos muestras.

***10.114** Consulte el Ejercicio 10.113. Suponga que otra muestra aleatoria independiente de tamaño n_3 se selecciona de una tercera población normal con media μ_3 y varianza σ^2. Encuentre la prueba de razón de probabilidades para probar $H_0: \mu_1 = \mu_2 = \mu_3$ contra la alternativa de que hay al menos una desigualdad. Demuestre que esta prueba es equivalente a una prueba F exacta.

10.12 Resumen

En los Capítulos 8, 9 y 10 hemos presentado los conceptos básicos relacionados con dos métodos para hacer inferencias: estimación y pruebas de hipótesis. Filosóficamente, la estimación (Capítulos 8 y 9) se concentra en esta pregunta: ¿cuál es el valor numérico de un parámetro θ? En contraste, una prueba de una hipótesis trata de contestar esta pregunta: ¿hay suficiente evidencia para apoyar la hipótesis alternativa? A menudo, el método inferencial que se emplea para una situación dada depende de cómo el experimentador prefiere expresar su inferencia. A veces esta decisión se saca del bolsillo; es decir, la pregunta práctica claramente implica que se use una estimación o un procedimiento de hipótesis. Por ejemplo, la aceptación o rechazo de los suministros entrantes o de los productos de salida en un proceso de manufactura claramente requiere una decisión o una prueba estadística. Hemos visto que existe dualidad entre estos dos procedimientos de hacer inferencias. Un intervalo de confianza bilateral con coeficiente de confianza $1 - \alpha$ puede verse como el conjunto de todos los valores de θ_0 que son valores "aceptables" de hipótesis nula para θ si usamos una prueba de nivel α bilateral. Del mismo modo, una prueba de nivel α de dos lados para $H_0: \theta = \theta_0$ se puede poner en práctica al construir un intervalo de confianza bilateral (con coeficiente de confianza $1 - \alpha$) y rechazar H_0 si el valor θ_0 *cae fuera* del intervalo de confianza.

Asociadas con ambos métodos para hacer inferencias hay medidas de su bondad. Así, el ancho esperado de un intervalo de confianza y el coeficiente de confianza miden la bondad del procedimiento de estimación. Del mismo modo, la bondad de una prueba estadística es medida por las probabilidades α y β de cometer errores tipo I y tipo II. Estas medidas de bondad hacen posible que comparemos una prueba estadística con otra y desarrollemos una teoría para adquirir pruebas estadísticas con propiedades deseables. La capacidad de evaluar la bondad de una inferencia es una de las principales aportaciones de la estadística al análisis de datos experimentales. ¿Cuál es el valor de una inferencia si no tenemos medida de su validez?

En este capítulo hemos investigado los elementos de una prueba estadística y explicado la forma en que trabaja una prueba. Algunas pruebas útiles se dan para demostrar cómo se pueden usar en situaciones prácticas, además veremos otras interesantes aplicaciones en los siguientes capítulos.

Muchos de los procedimientos de prueba desarrollados en este capítulo se presentaron desde un punto de vista intuitivo, pero también hemos ejemplificado el uso del lema de Neyman−Pearson para obtener procedimientos más potentes para probar una hipótesis nula simple contra una hipótesis alternativa simple. Además, hemos visto la forma en que a veces se puede usar el método de Neyman−Pearson para hallar pruebas uniformemente más potentes para hipótesis compuestas nulas y alternativas si la distribución básica se especifica, excepto por el valor de un parámetro individual. La razón de probabilidades produce un método general para desarrollar una prueba estadística. Las pruebas de razón de verosimilitudes se pueden efectuar ya sea que existan o no parámetros de ruido. En general, las pruebas de razón de probabilidades poseen propiedades deseables. Los procedimientos de Neyman−Pearson y de razón de verosimilitudes requieren que la distribución de la(s) población(es) mostrada(s) sea conocida, excepto para los valores de algunos parámetros. De otro modo, las funciones de probabilidad no pueden ser determinadas y los métodos no se pueden aplicar.

Bibliografía y lecturas adicionales

Casella, G., and R. L. Berger. 2002. *Statistical Inference,* 2d ed. Pacific Grove, Calif.: Duxbury.

Cramer, H. 1963. *Mathematical Methods of Statistics.* Princeton, N.J.: Princeton University Press.

Hoel, P.G. 1984. *Introduction to Mathematical Statistics,* 5th ed. New York: Wiley.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics,* 6th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Lehmann, E. L., and J. P. Romano. 2006. *Testing Statistical Hypotheses,* 3d ed. New York: Springer.

Miller, I., and M. Miller. 2003. *John E. Freund's Mathematical Statistics with Applications,* 7th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics,* 3d ed. New York: McGraw-Hill.

Ejercicios complementarios

10.115 Verdadero o falso.

a Si el valor p para una prueba es .036, la hipótesis nula puede ser rechazada en el nivel de significancia $\alpha = .05$.

b En una prueba formal de hipótesis, α es la probabilidad de que la hipótesis nula sea incorrecta.

c Si el valor p es muy pequeño para una prueba que compara dos medias poblacionales, la diferencia entre las medias debe ser grande.

d La potencia(θ^*) es la probabilidad de que la hipótesis nula sea rechazada cuando $\theta = \theta^*$.

e La potencia(θ) se calcula siempre suponiendo que la hipótesis nula es verdadera.

f Si .01 < valor p < .025, la hipótesis nula siempre puede ser rechazada en el nivel de significancia $\alpha = .02$.

g Suponga que una prueba es una prueba nivel α uniformemente más poderosa respecto al valor de un parámetro θ. Si θ_a es un valor en la hipótesis alternativa, $\beta(\theta_a)$ podría ser menor para alguna otra prueba nivel α.

h Cuando se desarrolla una prueba de razón de probabilidad, es posible que $L(\hat{\Omega}_0) > L(\hat{\Omega})$.

i $-2\ln(\lambda)$ es siempre positivo.

10.116 Consulte el Ejercicio 10.6. Encuentre la potencia(p), para $p = .2, .3, .4, .5, .6, .7$ y $.8$ y trace un dibujo aproximado de la función de potencia.

10.117 Lord Rayleigh fue uno de los primeros científicos en estudiar la densidad del nitrógeno. En sus estudios observó algo peculiar. Las densidades del nitrógeno producido a partir de compuestos químicos tendían a ser menores que las densidades del nitrógeno producido del aire. Las mediciones de Lord Rayleigh[18] se dan en la siguiente tabla. Estas mediciones corresponden a la masa de nitrógeno que llena un frasco de volumen especificado a temperatura y presión especificadas.

Compuesto químico	Atmósfera
2.30143	2.31017
2.29890	2.30986
2.29816	2.31010
2.30182	2.31001
2.29869	2.31024
2.29940	2.31010
2.29849	2.31028
2.29889	2.31163
2.30074	2.30956
2.30054	

a Para las mediciones del compuesto químico, $\overline{y} = 2.29971$ y $s = .001310$; para las mediciones de la atmósfera, $\overline{y} = 2.310217$ y $s = .000574$. ¿Hay suficiente evidencia para indicar una diferencia en la masa media de nitrógeno por frasco para compuestos químicos y el aire? ¿Qué se puede decir acerca del valor p asociado con su prueba?

b Encuentre un intervalo de confianza de 95% para la diferencia en masa media de nitrógeno por frasco para compuestos químicos y el aire.

c Con base en su respuesta al inciso b, en el nivel de significancia $\alpha = .05$, ¿hay suficiente evidencia para indicar una diferencia en la masa media del nitrógeno por frasco para mediciones de compuestos químicos y el aire?

d ¿Hay algún conflicto entre sus conclusiones en los incisos a y b? Aun cuando la diferencia en estas masas medias de nitrógeno es pequeña, Lord Rayleigh destacó esta diferencia en lugar de pasarla por alto y esto llevó al descubrimiento de los gases inertes en la atmósfera.

10.118 El efecto del consumo de alcohol parece ser mucho mayor a altitudes superiores. Para probar esta teoría, un científico seleccionó al azar 12 individuos y los dividió en dos grupos de 6 cada uno. Un grupo fue transportado a una altitud de 12,000 pies y cada miembro del grupo ingirió 100 centímetros cúbicos (cm^3) de alcohol. Los miembros del segundo grupo fueron llevados al nivel del mar y recibieron la misma cantidad de alcohol. Después de dos horas, se midió la cantidad de alcohol en el torrente sanguíneo de cada sujeto (mediciones en gramos/100 cm^3). Los datos aparecen en la siguiente tabla. ¿Hay suficiente evidencia para indicar que la retención de alcohol es mayor a 12,000 pies que al nivel del mar? Pruebe en el nivel de significancia $\alpha = .10$.

Nivel del mar	12,000 pies
.07	.13
.10	.17
.09	.15
.12	.14
.09	.10
.13	.14

18. *Fuente*: *Proceedings*, *Royal Society* (Londres) 55 (1894): 340–344.

10.119 En la actualidad, 20% de los clientes potenciales compran jabón de la marca A. Para incrementar ventas, la compañía realizó una intensa campaña de publicidad. Al final de la campaña, una muestra de 400 clientes potenciales fue entrevistada para determinar si la campaña tuvo éxito.

a Exprese H_0 y H_a en términos de p, la probabilidad de que un cliente prefiera el jabón de la marca A.

b La compañía decide concluir que la campaña de publicidad fue un éxito si al menos 92 de los 400 clientes entrevistados prefería la marca A. Encuentre α. (Use la aproximación normal a la distribución binomial para evaluar la probabilidad deseada.)

10.120 En el pasado, una planta química ha producido un promedio de 1100 libras de productos químicos al día. Los registros del año pasado, con base en 260 días de operación, muestran lo siguiente:

$$\overline{y} = 1060 \text{ libras/día}, \qquad s = 340 \text{ libras/día}.$$

Deseamos probar si el promedio de producción diaria bajó sensiblemente el año pasado.

a Formule las hipótesis nula y alternativa apropiadas.

b Si se utiliza Z como estadístico de prueba, determine la región de rechazo correspondiente a un nivel de significancia de $\alpha = .05$.

c ¿Los datos aportan suficiente evidencia para indicar una caída en el promedio de producción diaria?

10.121 Se comparó la capacidad de frenado de dos tipos de automóvil y se probaron muestras aleatorias de 64 automóviles para cada tipo. Las mediciones registradas fueron las distancias requeridas para detenerse cuando se aplicaban los frenos a una velocidad de 40 millas por hora. Las medias muestrales y varianzas calculadas fueron las siguientes:

$$\overline{y}_1 = 118, \quad \overline{y}_2 = 109,$$
$$s_1^2 = 102, \quad s_2^2 = 87.$$

¿Los datos aportan suficiente evidencia para indicar una diferencia en las distancias medias de parada de los dos tipos de automóviles? Determine el nivel de significancia alcanzado.

10.122 La estabilidad de las mediciones de las características de un producto manufacturado es importante para mantener la calidad del mismo. De hecho, a veces es mejor obtener una pequeña variación en el valor medido de alguna característica importante de un producto y tener la media del proceso ligeramente fuera del objetivo, que obtener una amplia variación con un valor medio que se ajuste perfectamente a los requerimientos. Esta última situación puede producir un porcentaje más alto de piezas defectuosas que la primera. Un fabricante sospechaba que una de sus líneas de producción estaba produciendo focos con una alta variación en su duración. Para probar su teoría, comparó las duraciones de $n = 50$ focos muestreados al azar tomados de la línea sospechosa y $n = 50$ de una línea que parecía estar bajo control. Las medias muestrales y varianzas para las dos muestras fueron como se ve en la siguiente tabla.

Línea sospechosa	Línea bajo control
$\overline{y}_1 = 1{,}520$	$\overline{y}_2 = 1{,}476$
$s_1^2 = 92{,}000$	$s_2^2 = 37{,}000$

a ¿Los datos aportan suficiente evidencia para indicar que los focos producidos por la línea sospechosa poseen una duración mayor que los producidos por la línea que se supone está bajo control? Use $\alpha = .05$.

b Encuentre el nivel aproximado de significancia observado para la prueba e interprete su valor.

10.123 Un fabricante de productos farmacéuticos compra un material particular a dos diferentes proveedores. El nivel medio de impurezas de la materia prima es aproximadamente igual para ambos proveedores, pero el fabricante está preocupado por la variabilidad de las impurezas de un embarque a otro. Si el nivel de impurezas tiende a variar de manera excesiva para una fuente de suministro, podría afectar la calidad del producto farmacéutico. Para comparar la variación en el porcentaje de impurezas para los dos proveedores, el fabricante selecciona diez embarques de cada uno de ellos y mide el porcentaje de impurezas de la materia prima para cada embarque. Las medias muestrales y varianzas se observan en la siguiente tabla.

Proveedor A	Proveedor B
$\overline{y}_1 = 1.89$	$\overline{y}_2 = 1.85$
$s_1^2 = .273$	$s_2^2 = .094$
$n_1 = 10$	$n_2 = 10$

a ¿Los datos dan suficiente evidencia para indicar una diferencia en la variabilidad de los niveles de impureza de los embarques para los dos proveedores? Pruebe usando $\alpha = .10$. Con base en los resultados de la prueba, ¿qué recomendación haría usted al fabricante farmacéutico?

b Encuentre un intervalo de confianza de 90% para σ_B^2 e interprete sus resultados.

10.124 Los datos de la siguiente tabla muestran lecturas en pies-libras de la resistencia al impacto de dos clases de material de empaque, tipo A y tipo B. Determine si la información sugiere una diferencia en la resistencia media entre las dos clases de material. Realice la prueba con un nivel de significancia $\alpha = .10$.

A	B
1.25	.89
1.16	1.01
1.33	.97
1.15	.95
1.23	.94
1.20	1.02
1.32	.98
1.28	1.06
1.21	.98
$\sum y_i = 11.13$	$\sum y_i = 8.80$
$\overline{y} = 1.237$	$\overline{y} = .978$
$\sum y_i^2 = 13.7973$	$\sum y_i^2 = 8.6240$

10.125 ¿Qué eficiencia de combustión debe esperar el propietario de una casa de un horno de petróleo? La EPA indica que 80% o más es excelente, 75% a 79% es buena, 70% a 74% es regular y debajo de 70% es mala. Un contratista de sistemas de calefacción doméstica, que vende dos marcas de calentadores de petróleo (llamémosles A y B) decidió comparar sus eficiencias medias al analizar las eficiencias de 8 calentadores del tipo A y 6 del tipo B. Los porcentajes de eficiencia resultantes para los 14 calentadores se muestran en la tabla siguiente.

Tipo A	Tipo B
72	78
78	76
73	81
69	74
75	82
74	75
69	
75	

a ¿Los datos proporcionan suficiente evidencia para indicar una diferencia en las eficiencias medias para las dos marcas de calentadores domésticos? Encuentre el valor p aproximado para la prueba e interprete su valor.

b Determine un intervalo de confianza de 90% para $(\mu_A - \mu_B)$ e interprete el resultado.

10.126 Suponga que $X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}$, y $W_1, W_2, \ldots, W_{n_3}$ son muestras aleatorias independientes de distribuciones normales con sus respectivas medias desconocidas μ_1, μ_2 y μ_3 y varianzas comunes $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$. Suponga que deseamos calcular una función lineal de las medias: $\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3$. Debido a que el estimador de máxima verosimilitud (MLE) de una función de parámetros es la función de los MLE de los parámetros, el MLE de θ es $\hat{\theta} = a_1\overline{X} + a_2\overline{Y} + a_3\overline{W}$.

a ¿Cuál es el error estándar del estimador $\hat{\theta}$?

b ¿Cuál es la distribución del estimador $\hat{\theta}$?

c Si las varianzas muestrales están dadas por S_1^2, S_2^2 y S_3^2, respectivamente, considere

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{n_1 + n_2 + n_3 - 3}.$$

i ¿Cuál es la distribución de $(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2$?

ii ¿Cuál es la distribución de

$$T = \frac{\hat{\theta} - \theta}{S_p\sqrt{\dfrac{a_1^2}{n_1} + \dfrac{a_2^2}{n_2} + \dfrac{a_3^2}{n_3}}}?$$

d Dé un intervalo de confianza para θ con coeficiente de confianza $1 - \alpha$.

e Desarrolle una prueba para $H_0: \theta = \theta_0$ contra $H_a: \theta \neq \theta_0$.

10.127 Un comerciante piensa que su utilidad semanal es una función de tres variables: ventas al menudeo (denotadas por X), ventas al mayoreo (denotadas por Y) y gastos generales (denotados por W). Las variables X, Y y W son consideradas como variables aleatorias distribuidas normalmente, independientes, con medias μ_1, μ_2 y μ_3 y varianzas σ^2, $a\sigma^2$ y $b\sigma^2$, respectivamente, para constantes conocidas a y b, pero σ^2 desconocida. La utilidad semanal esperada del comerciante es $\mu_1 + \mu_2 - \mu_3$. Si el comerciante ha hecho observaciones independientes de X, Y y W durante las últimas n semanas, construya una prueba de $H_0: \mu_1 + \mu_2 - \mu_3 = k$, contra la alternativa $H_a: \mu_1 + \mu_2 - \mu_3 \neq k$, para una constante dada k. Puede especificar $\alpha = .05$.

10.128 Un examen de lectura se aplica a alumnos de sexto año en tres grandes escuelas primarias. Se considera que las calificaciones del examen en cada escuela tienen distribuciones normales con medias desconocidas μ_1, μ_2 y μ_3, respectivamente, y varianza común desconocida $\sigma^2(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2)$. Usando los datos de la tabla siguiente en muestras aleatorias independientes de cada escuela, pruebe si existe evidencia de una diferencia entre μ_1 y μ_2. Use $\alpha = .05$.

Escuela I	Escuela II	Escuela III
$n_1 = 10$	$n_2 = 10$	$n_3 = 10$
$\sum x_i^2 = 36{,}950$	$\sum y_i^2 = 25{,}850$	$\sum w_i^2 = 49{,}900$
$\overline{x} = 60$	$\overline{y} = 50$	$\overline{w} = 70$

***10.129** Suponga que $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de la función de densidad de probabilidad dada por

$$f(y \mid \theta_1, \theta_2) = \begin{cases} \left(\dfrac{1}{\theta_1}\right) e^{-(y-\theta_2)/\theta_1}, & y > \theta_2, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Determine la prueba de razón de verosimilitudes para probar $H_0 : \theta_1 = \theta_{1,0}$ contra $H_a : \theta_1 > \theta_{1,0}$ con θ_2 desconocida.

***10.130** Consulte el Ejercicio 10.129. Encuentre la prueba de razón de verosimilitudes probando $H_0 : \theta_2 = \theta_{2,0}$ contra $H_a : \theta_2 > \theta_{2,0}$ con θ_1 desconocida.

CAPÍTULO 11

Modelos lineales y estimación por mínimos cuadrados

11.1 Introducción

En el Capítulo 9 consideramos varios métodos para hallar estimadores de parámetros, incluyendo los métodos de momentos y máxima probabilidad, así como los basados en estadísticos suficientes. Otro método de estimación, el de mínimos cuadrados, es el tema de este capítulo.

En todas nuestras explicaciones anteriores acerca de la inferencia estadística supusimos que las variables aleatorias $Y_1, Y_2, \ldots, Y_n$, eran independientes y distribuidas idénticamente. Una implicación de esta suposición es que el valor esperado de Y_i, $E(Y_i)$, es constante (si existe). Esto es, $E(Y_i) = \mu$ no depende del valor de ninguna otra variable. Obviamente, esta suposición no es válida en muchos problemas inferenciales. Por ejemplo, la distancia media de frenado para cierto tipo de automóvil dependerá de la rapidez a la que el automóvil se esté moviendo; la potencia media de un antibiótico depende del tiempo que éste haya estado almacenado; la cantidad media de alargamiento observada en una aleación de metal depende de la fuerza aplicada y la temperatura de la aleación. En este capítulo emprendemos un estudio de procedimientos inferenciales que se pueden usar cuando una variable aleatoria Y, llamada *variable dependiente*, tiene una media que es función de una o más variables no aleatorias x_1, $x_2, \ldots, x_k$, llamadas *variables independientes*. (En este contexto, los términos *independiente* y *dependiente* se usan en su sentido matemático. No hay relación con el concepto probabilístico de variables aleatorias independientes.)

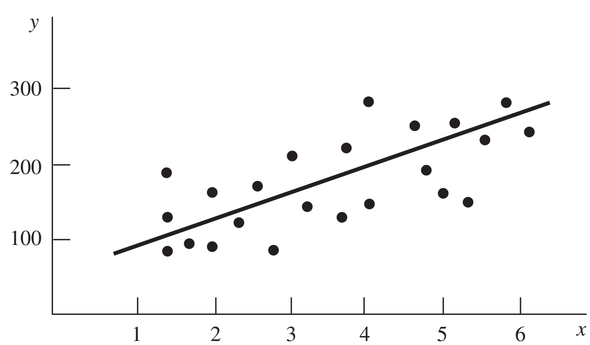
Muchos tipos diferentes de funciones matemáticas se pueden usar para modelar una respuesta que sea una función de una o más variables independientes. Éstas se pueden clasificar en dos categorías: modelos determinísticos y probabilísticos. Por ejemplo, suponga que y y x están relacionadas de acuerdo con la ecuación

$$y = \beta_0 + \beta_1 x,$$

donde β_0 y β_1 son parámetros desconocidos. Este modelo se llama modelo matemático *determinístico* porque no toma en cuenta ningún error para predecir o pronosticar y como función de x. Este modelo implica que y siempre tome el valor $\beta_0 + \beta_1(5.5)$ siempre que $x = 5.5$.

Suponga que recolectamos una muestra de n valores de y correspondientes a n ajustes diferentes de la variable independiente x y que una gráfica de los datos es como se ilustra en la Figura 11.1. Está claro en la figura que el valor esperado de Y puede aumentar como función lineal de x, pero un modelo determinístico está lejos de una descripción adecuada de

FIGURA **11.1** Gráfica de datos



la realidad. Experimentos repetidos cuando $x = 5.5$ darían valores de Y que variarán de manera aleatoria, lo cual nos dice que el modelo determinístico no es una representación exacta de la relación entre las dos variables. Además, si el modelo se usara para predecir Y cuando $x = 5.5$, la predicción estaría sujeta a algún error desconocido y esto, por supuesto, nos lleva al uso de métodos estadísticos. Predecir Y para un valor dado de x es un proceso inferencial. Si la predicción ha de ser de utilidad en la vida real, es necesario que estemos en posibilidad de evaluar la verosimilitud de observar errores de predicción de varias magnitudes.

En contraste con el modelo determinístico, los expertos en estadística usan modelos *probabilísticos*. Por ejemplo, podríamos representar las respuestas de la Figura 11.1 por medio del modelo

$$E(Y) = \beta_0 + \beta_1 x$$

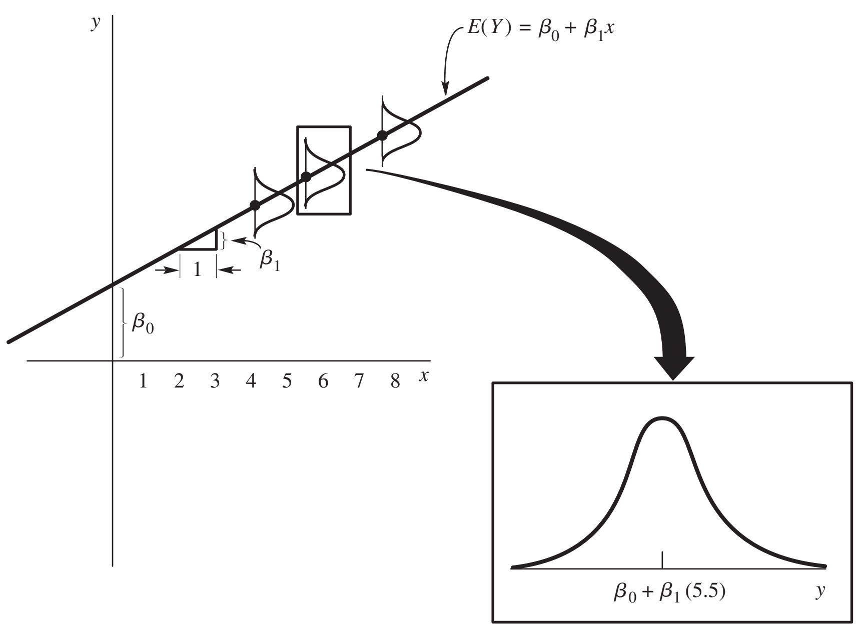
o bien, lo que es equivalente,

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

donde ε es una variable aleatoria que tiene una distribución de probabilidad específica con media 0. Consideremos a Y como la suma de un componente determinístico $E(Y)$ y un componente aleatorio ε. Este modelo toma en cuenta el comportamiento aleatorio de Y exhibido en la Figura 11.1 y da una descripción más precisa de la realidad que el modelo determinístico. Además, las propiedades del error de predicción para Y se pueden obtener en muchos de los modelos probabilísticos.

La Figura 11.2 muestra una representación gráfica del modelo probabilístico $Y = \beta_0 + \beta_1 x + \varepsilon$. Cuando $x = 5.5$, hay una *población* de posibles valores de Y. La distribución de esta población está indicada en la parte principal de la gráfica y está centrada en la recta $E(Y) = \beta_0 + \beta_1 x$ en el punto $x = 5.5$. Esta población tiene una distribución con media $\beta_0 + \beta_1(5.5)$ y varianza σ^2, como se muestra en la versión amplificada de la distribución que está encerrada

FIGURA **11.2** Gráfica del modelo probabilístico $Y = \beta_0 + \beta_1 x + \varepsilon$



en el recuadro de la Figura 11.2. Cuando $x = 7$, hay otra *población* de posibles valores de Y. La distribución de esta población tiene la misma forma que la distribución de valores Y cuando $x = 5.5$ y tiene la misma varianza σ^2, pero cuando $x = 7$, la distribución de Y tiene media $\beta_0 + \beta_1(7)$. Lo mismo se cumple para cada posible valor de la variable independiente x. Esto es, en un modelo de regresión, existe una *población separada* de valores de respuesta para cada posible ajuste de la(s) variable(s) independiente(s). Todas estas poblaciones tienen la misma varianza y la forma de las distribuciones de las poblaciones son iguales (véase la Figura 11.2); no obstante, la media de cada población depende, mediante el modelo de regresión, del ajuste de la(s) variable(s) independiente(s). Existen textos científicos y matemáticos llenos de modelos determinísticos de la realidad. De hecho, muchas de las funciones matemáticas que aparecen en libros de cálculo y física son modelos matemáticos determinísticos de la naturaleza. Por ejemplo, la ley de Newton que relaciona la fuerza de un cuerpo en movimiento con su masa y aceleración,

$$F = ma,$$

es un modelo determinístico que, para fines prácticos, predice con poco error. En contraste, otros modelos —por ejemplo las funciones representadas gráficamente en publicaciones y textos científicos— con frecuencia son deficientes. Se ha restado énfasis a la dispersión de puntos que darían evidencia gráfica de sus inadecuaciones, semejantes al comportamiento aleatorio de los puntos de la Figura 11.1, lo cual lleva a los científicos novatos a aceptar las correspondientes "leyes" y teorías como una descripción exacta de la naturaleza.

Si se pueden usar modelos determinísticos para predecir con error insignificante, para todos los fines prácticos, los usamos. Si no es así, buscamos un modelo probabilístico que no será una caracterización exacta de la naturaleza pero que hace posible evaluar la validez de nuestras inferencias.

11.2 Modelos estadísticos lineales

Aun cuando se puede usar un número infinito de funciones diferentes para modelar el valor medio de la variable de respuesta Y como función de una o más variables independientes, nos concentraremos en un conjunto de modelos llamados *modelos estadísticos lineales*. Si Y es la variable de respuesta y x es una sola variable independiente, puede ser razonable en algunas situaciones usar el modelo $E(Y)= \beta_0 + \beta_1 x$ para los valores desconocidos de parámetros β_0 y β_1. Observe que, en este modelo, $E(Y)$ es una función lineal de x (para β_0 y β_1 dadas) y también una función lineal de β_0 y β_1 [porque $E(Y)= c\beta_0 + d\beta_1$ con $c = 1$ y $d = x$]. En el modelo $E(Y)= \beta_0 + \beta_1 x^2$, $E(Y)$ no es una función lineal de x, pero *es* una función lineal de β_0 y β_1 [porque $E(Y)= c\beta_0 + d\beta_1$ con $c = 1$ y $d = x^2$]. Cuando decimos que tenemos un modelo estadístico lineal para Y, queremos decir que $E(Y)$ es una función lineal de los parámetros desconocidos β_0 y β_1 y *no* necesariamente una función lineal de x. Entonces, $Y = \beta_0 + \beta_1 (\ln x) + \varepsilon$ es un modelo lineal (porque $\ln x$ toma valores conocidos para cada valor fijo de x).

Si el modelo relaciona $E(Y)$ como una función lineal de β_0 y β_1 únicamente, el modelo recibe el nombre de modelo de regresión lineal *simple*. Si más de una variable independiente, por ejemplo $x_1, x_2, \ldots, x_k$, son de interés y modelamos $E(Y)$ con

$$E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k,$$

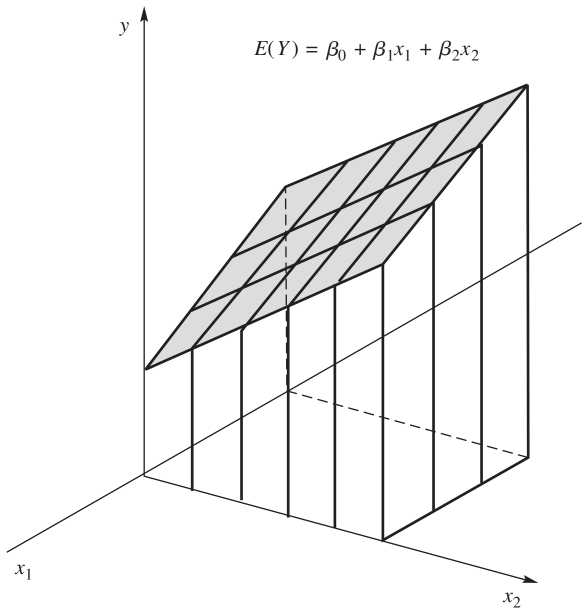


FIGURA **11.3** Gráfica de $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

el modelo se denomina modelo de regresión lineal *múltiple*. Como $x_1, x_2, \ldots, x_k$ son consideradas como variables con valores conocidos, se supone que están medidas sin error en un experimento. Por ejemplo, si usted piensa que el rendimiento medio $E(Y)$ es una función de la variable t, la temperatura de un proceso químico, podría hacer $x_1 = t$ y $x_2 = e^t$ y usar el modelo $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ o bien, de manera equivalente $E(Y) = \beta_0 + \beta_1 t + \beta_2 e^t$. O bien, si $E(Y)$ es una función de dos variables x_1 y x_2, podría escoger una aproximación plana a la verdadera respuesta media usando el modelo lineal $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Entonces, $E(Y)$ es una función lineal de β_0, β_1 y β_2 y representa un plano en el espacio y, x_1, x_2 (vea la Figura 11.3). Del mismo modo,

$$E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

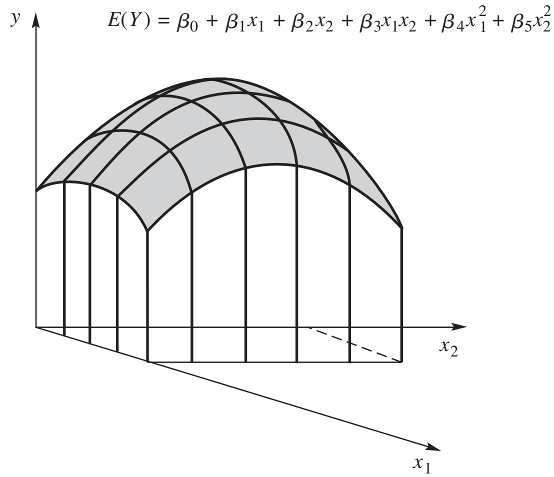
es un modelo estadístico lineal, donde $E(Y)$ es una función con polinomios de segundo orden de la variable independiente x, con $x_1 = x$ y $x_2 = x^2$. Este modelo sería apropiado para una respuesta que traza un segmento de una parábola sobre la región experimental.

El porcentaje de agua esperado $E(Y)$ que contiene el papel durante su manufactura podría estar representado por una función de segundo orden de la temperatura del secador, x_1, y la rapidez de la máquina que fabrica el papel, x_2. Entonces,

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2,$$

donde $\beta_0, \beta_1, \ldots, \beta_5$ son parámetros desconocidos en el modelo. Geométricamente, $E(Y)$ describe una superficie (cónica) de segundo orden sobre el plano x_1, x_2 (vea la Figura 11.4).

FIGURA **11.4**
Gráfica de $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$



DEFINICIÓN **11.1**

Un modelo *estadístico lineal* que relaciona una respuesta aleatoria Y con un conjunto de variables independientes $x_1, x_2, \ldots, x_k$ es de la forma

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon,$$

donde $\beta_0, \beta_1, \ldots, \beta_k$ son parámetros desconocidos, ε es una variable aleatoria y las variables $x_1, x_2, \ldots, x_k$ toman valores conocidos. Supondremos que $E(\varepsilon) = 0$ y por tanto que

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

Consideremos la interpretación física del modelo lineal Y. Decimos que Y es igual a un valor esperado $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ (una función de las variables independientes $x_1, x_2, \ldots, x_k$), más un error aleatorio ε. Desde un punto de vista práctico, ε reconoce nuestra incapacidad para dar un modelo exacto por naturaleza. En experimentación repetida, Y varía alrededor de $E(Y)$ de un modo aleatorio porque no hemos incluido en nuestro modelo toda la gran cantidad de variables que pueden afectar a Y. Por fortuna, muchas veces el efecto neto de estas variables no medidas y con mucha frecuencia desconocidas, es hacer que Y varíe de manera que puedan ser aproximadas adecuadamente mediante una suposición de comportamiento aleatorio.

En este capítulo usamos el *método de mínimos cuadrados* para obtener estimadores para los parámetros $\beta_0, \beta_1, \ldots, \beta_k$ en un modelo de regresión lineal. En muchas aplicaciones uno o más de estos parámetros tendrán interpretaciones significativas, razón por la cual desarrollamos métodos inferenciales para un parámetro β individual y para conjuntos de parámetros β. Si estimamos los parámetros $\beta_0, \beta_1, \ldots, \beta_5$ del modelo que expresa el porcentaje esperado $E(Y)$ de agua en papel como polinomio de segundo orden en x_1 (la temperatura del secador) y x_2 (la rapidez del secador), podremos desarrollar métodos para estimar y formar intervalos de confianza para el valor de $E(Y)$

cuando x_1 y x_2 tomen valores específicos. Del mismo modo, podemos desarrollar métodos para predecir un valor futuro de Y cuando las variables independientes tomen valores de interés práctico. Las Secciones 11.3 a 11.9 se concentran en el modelo de regresión lineal *simple*, mientras que las últimas secciones se refieren a modelos de regresión lineal *múltiple*.

11.3 Método de mínimos cuadrados

Un procedimiento para estimar los parámetros de cualquier modelo lineal, el método de mínimos cuadrados, se puede ilustrar con sólo ajustar una recta a un conjunto de puntos. Suponga que deseamos ajustar el modelo

$$E(Y) = \beta_0 + \beta_1 x$$

al conjunto de puntos que se muestra en la Figura 11.5. [La variable independiente x podría ser w^2 o $(w)^{1/2}$ o $\ln w$, etc., para alguna otra variable independiente w.] Esto es, postulamos que $Y = \beta_0 + \beta_1 x + \varepsilon$, donde ε tiene alguna distribución de probabilidad con $E(\varepsilon) = 0$. Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores de los parámetros β_0 y β_1, entonces $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ es claramente un estimador de $E(Y)$.

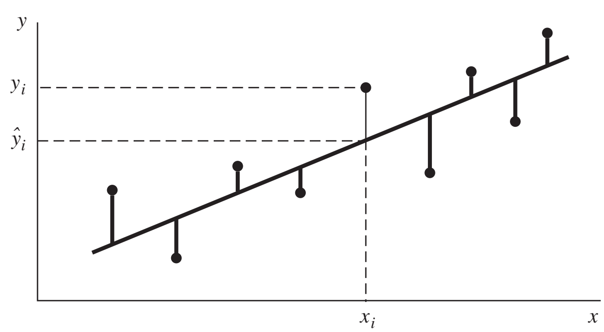
El procedimiento de mínimos cuadrados para ajustar una recta que pase por un conjunto de n puntos es semejante al método que podríamos usar si ajustamos una recta a simple vista; esto es, deseamos que las diferencias entre los valores observados y los puntos correspondientes en la recta ajustada sean "pequeñas" en un sentido general. Una forma cómoda de lograr esto y que proporciona estimadores con buenas propiedades, es minimizar la suma de cuadrados de las desviaciones verticales a partir de la recta ajustada (vea las desviaciones indicadas en la Figura 11.5). Entonces, si

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

es el valor pronosticado del i-ésimo valor y (cuando $x = x_i$), entonces la desviación (a veces llamada *error*) del valor observado de y_i a partir de $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ es la diferencia $y_i - \hat{y}_i$ y la suma de los cuadrados de las desviaciones a minimizar es

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

FIGURA **11.5** Ajuste de una recta que pasa por un conjunto de puntos



La cantidad SSE también recibe el nombre de *suma de cuadrados del error* por razones que más adelante se harán evidentes.

Si la SSE tiene un mínimo, ocurrirá para valores de β_0 y β_1 que satisfagan las ecuaciones, $\partial \text{SSE}/\partial \hat{\beta}_0 = 0$ y $\partial \text{SSE}/\partial \hat{\beta}_1 = 0$. Tomando las derivadas parciales de la SSE con respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$ e igualando a cero, obtenemos

$$\frac{\partial \text{SSE}}{\partial \hat{\beta}_0} = \frac{\partial \left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial \hat{\beta}_0} = -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]$$

$$= -2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) = 0$$

y

$$\frac{\partial \text{SSE}}{\partial \hat{\beta}_1} = \frac{\partial \left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial \hat{\beta}_1} = -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]x_i$$

$$= -2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2\right) = 0.$$

Las ecuaciones $\partial \text{SSE}/\hat{\beta}_0 = 0$ y $\partial \text{SSE}/\hat{\beta}_1 = 0$ se denominan *ecuaciones de mínimos cuadrados* para estimar los parámetros de una recta.

Las ecuaciones de mínimos cuadrados son lineales en $\hat{\beta}_0$ y $\hat{\beta}_1$ y por tanto pueden resolverse simultáneamente. Usted puede verificar que las soluciones son

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2},$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

Además, se puede demostrar que la solución simultánea para las dos ecuaciones de mínimos cuadrados da valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimizan la SSE. Dejamos esto para que lo compruebe.

Las expresiones

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad \text{y} \quad \sum_{i=1}^{n}(x_i - \overline{x})^2$$

que se usan para calcular $\hat{\beta}_1$ se encuentran a menudo en el desarrollo de modelos de regresión lineal simple. La primera de éstas se calcula al sumar productos de valores x menos su media y valores y menos su media. En todas nuestras exposiciones siguientes denotaremos esta cantidad por S_{xy}. Del mismo modo, denotaremos la segunda cantidad por S_{xx} porque se calcula al sumar productos que contienen únicamente los valores x.

Estimadores de mínimos cuadrados para el modelo de regresión lineal simple

1. $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$, donde $S_{xy} = \displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ y $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$.
2. $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$.

Ilustramos el uso de las ecuaciones anteriores con un ejemplo sencillo.

EJEMPLO **11.1** Use el método de mínimos cuadrados para ajustar una recta a los $n = 5$ puntos de datos dados en la Tabla 11.1

Tabla 11.1 Datos para el Ejemplo 11.1

x	y
−2	0
−1	0
0	1
1	1
2	3

Solución Comenzamos el cálculo de las estimaciones de mínimos cuadrados para la pendiente y puntos de intersección de la recta ajustada construyendo la Tabla 11.2. Usando los resultados de la tabla obtenemos

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\displaystyle\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\displaystyle\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{7 - \frac{1}{5}(0)(5)}{10 - \frac{1}{5}(0)^2} = .7,$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = \frac{5}{5} - (.7)(0) = 1,$$

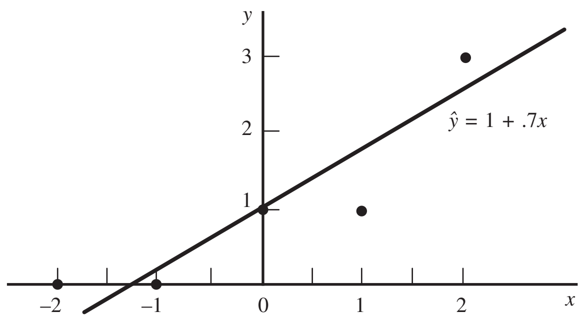
y la recta ajustada es

$$\hat{y} = 1 + .7x.$$

Tabla 11.2 Cálculos para determinar los coeficientes

x_i	y_i	$x_i y_i$	x_i^2
−2	0	0	4
−1	0	0	1
0	1	0	0
1	1	1	1
2	3	6	4
$\sum_{i=1}^{n} x_i = 0$	$\sum_{i=1}^{n} y_i = 5$	$\sum_{i=1}^{n} x_i y_i = 7$	$\sum_{i=1}^{n} x_i^2 = 10$

FIGURA **11.6** Gráfica de puntos y recta de mínimos cuadrados para el Ejemplo 11.1



Los cinco puntos y la recta ajustada se muestran en la Figura 11.6. ■

En esta sección, hemos determinado los estimadores de mínimos cuadrados para los parámetros β_0 y β_1 del modelo $E(Y) = \beta_0 + \beta_1 x$. El sencillo ejemplo empleado aquí reaparecerá en futuras secciones para ilustrar otros cálculos. Ejercicios de naturaleza más realista se presentan al final de las secciones y se analizan dos ejemplos que comprenden datos provenientes de experimentos reales en la Sección 11.9. En la siguiente sección desarrollamos las propiedades estadísticas de los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$. Secciones subsiguientes están dedicadas a usar estos estimadores para varios fines inferenciales.

Ejercicios

11.1 Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son las estimaciones de mínimos cuadrados para la ordenada al origen y la pendiente en un modelo de regresión lineal sencilla, demuestre que la ecuación de mínimos cuadrados $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ siempre pasa por el punto $(\overline{x}, \overline{y})$. [*Sugerencia:* sustituya $\overline{x}$ por x en la ecuación de mínimos cuadrados y use el hecho de que $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.]

11.2 **Ejercicio Applet** ¿Cómo puede usted mejorar su comprensión de cómo trabaja realmente el método de mínimos cuadrados? Entre a la aplicación *Fitting a Line Using Least Squares* (en www.thomsonedu.com/statistics/wackerly). Los datos que aparecen en la primera gráfica son del Ejemplo 11.1.

a ¿Cuáles son la pendiente y la ordenada al origen de la recta horizontal azul? (Vea la ecuación sobre la gráfica.) ¿Cuál es la suma de los cuadrados de las desviaciones verticales entre los puntos sobre la recta horizontal y los valores observados de las y? ¿La recta horizontal se ajusta bien a los datos? Haga clic en el botón “Display/Hide Error Squares”. Observe que las áreas de los recuadros amarillos son iguales a los cuadrados de las desviaciones asociadas. ¿Cómo se compara la Suma de Cuadrados de los Errores (SSE) con la suma de las áreas de los recuadros amarillos?

b Haga clic en el botón “Display/Hide Error Squares” para que desaparezcan los recuadros amarillos. Ponga el cursor en el extremo derecho de la recta azul. Haga clic y mantenga presionado el botón del mouse y arrastre la recta para que la pendiente de la recta azul se haga negativa. ¿Qué observa acerca de las longitudes de las rectas verticales rojas? ¿La SSE aumenta o disminuye? ¿La recta con pendiente negativa parece ajustar bien los datos?

c Arrastre la recta para que la pendiente sea cercana a 0.8. ¿Qué ocurre cuando se acerca la pendiente a 0.7? ¿La SSE aumenta o disminuye? Cuando la recta azul se mueve en realidad está haciendo pivote alrededor de un punto fijo. ¿Cuáles son las coordenadas de ese punto pivote? ¿Las coordenadas del punto pivote son consistentes con el resultado que obtuvo en el Ejercicio 11.1?

d Arrastre la recta azul hasta que obtenga una recta que visualmente se ajuste bien a los datos. ¿Cuáles son la pendiente y la ordenada al origen de la recta que visualmente se ajusta a los datos? ¿Cuál es el valor de la SSE para la recta que visualmente se ajustó a los datos? Haga clic en el botón "Find Best Model" para obtener la recta de mínimos cuadrados. ¿Cómo se compara el valor de la SSE con la SSE asociada con la recta que usted visualmente ajustó a los datos? ¿Cómo se comparan la pendiente y la ordenada al origen de la recta que usted visualmente ajustó a los datos con la pendiente y la ordenada al origen de la recta de mínimos cuadrados?

11.3 Ajuste una recta a los cinco puntos de la tabla siguiente. Dé las estimaciones de β_0 y β_1. Localice los puntos y trace la recta ajustada como prueba de los cálculos.

y	3.0	2.0	1.0	1.0	0.5
x	−2.0	−1.0	0.0	1.0	2.0

11.4 Es frecuente que a los auditores se les exija comparar el valor auditado (o de lista) de un artículo de inventario contra el valor en libros. Si una empresa está llevando su inventario y libros actualizados, debería haber una fuerte relación lineal entre los valores auditados y en libros. Una empresa muestreó diez artículos de inventario y obtuvo los valores auditado y en libros que se dan en la tabla siguiente. Ajuste el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ a estos datos.

Artículo	Valor auditado (y_i)	Valor en libros (x_i)
1	9	10
2	14	12
3	7	9
4	29	27
5	45	47
6	109	112
7	40	36
8	238	241
9	60	59
10	170	167

a ¿Cuál es su estimación para el cambio esperado en valor auditado para un cambio de una unidad en el valor en libros?

b Si el valor en libros es $x = 100$, ¿qué usaría para estimar el valor auditado?

11.5 ¿Qué aspecto tenían los precios de vivienda en los "buenos y viejos tiempos"? La mediana de los precios de venta para casas nuevas unifamiliares se dan en la tabla siguiente, para los años 1972 a 1979.[1] Si con Y denotamos la mediana de los precios de venta y con x el año (usando enteros 1, 2, . . . , 8), ajuste el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$. ¿Qué se puede concluir de los resultados?

Año	Mediana de los precios de venta (×1000)
1972 (1)	\$27.6
1973 (2)	\$32.5
1974 (3)	\$35.9
1975 (4)	\$39.3
1976 (5)	\$44.2
1977 (6)	\$48.8
1978 (7)	\$55.7
1979 (8)	\$62.9

1. *Fuente*: adaptado de *Time*, 23 de julio de 1979, p. 67.

11.6 **Ejercicio Applet** Consulte los Ejercicios 11.2 y 11.5. Los datos del Ejercicio 11.5 aparecen en la gráfica bajo el encabezado “Another Example” en la aplicación *Fitting a Line Using Least Squares*. De nuevo, la recta azul horizontal que inicialmente aparece en la gráfica es una recta con pendiente 0.

a ¿Cuál es la ordenada al origen de la recta con pendiente 0? ¿Cuál es el valor de la SSE para la recta con pendiente 0?

b ¿Piensa usted que una recta con pendiente negativa se ajustará bien a los datos? Si la recta es arrastrada para producir una pendiente negativa, ¿la SSE aumenta o disminuye?

c Arrastre la recta para obtener una línea que visualmente se ajuste bien a los datos. ¿Cuál es la ecuación de la recta que obtuvo? ¿Cuál es el valor de la SSE? ¿Qué le ocurre a la SSE si la pendiente (y la ordenada al origen) de la recta se cambia desde la que usted ajusta visualmente?

d La recta que usted ajusta visualmente, ¿es la recta de mínimos cuadrados? Haga clic en el botón “Find Best Model” para obtener la recta con la SSE más pequeña. ¿Cómo se comparan la pendiente y la ordenada al origen de la recta de mínimos cuadrados con la pendiente y la ordenada al origen de la recta que usted visualmente ajustó en el inciso c? ¿Cómo se comparan las sumas de cuadrados del error (SSE)?

e Consulte el inciso a. ¿Cuál es la coordenada y del punto alrededor del cual gira la recta azul?

f Haga clic en el botón “Display/Hide Error Squares”. ¿Qué observa usted sobre el tamaño de los recuadros amarillos que aparecen en la gráfica? ¿Cuál es la suma de las áreas de los recuadros amarillos?

11.7 **Ejercicio Applet** Mueva la parte de la aplicación marcada como “Curvilinear Relationship” asociada con la aplicación *Fitting a Line Using Least Squares.*

a ¿Le parece que una recta dará un buen ajuste a los datos en la gráfica? ¿Parece que es probable que haya *alguna* relación funcional entre $E(Y)$ y x?

b ¿Hay alguna recta que se ajuste mejor a los datos que aquella con pendiente 0?

c Si usted ajusta una recta a un conjunto de datos y obtiene que la recta de mejor ajuste tiene pendiente 0, ¿significa eso que no hay relación funcional entre $E(Y)$ y la variable independiente? ¿Por qué?

11.8 Experimentos de laboratorio diseñados para medir valores de la LC50 (concentración letal que mata 50% de la especie a prueba), para el efecto de ciertos tóxicos en peces, son realizados siguiendo dos métodos. Uno consiste en circular agua continuamente por tanques de laboratorio y el otro mantiene agua en condiciones estáticas. Con el fin de establecer criterios de tóxicos, la Environmental Protection Agency (EPA) desea ajustar todos los resultados a la condición de circulación. Entonces, se hace necesario un modelo para relacionar los dos tipos de observaciones. Las observaciones de tóxicos examinados en condiciones estática y de flujo dieron los datos de la tabla siguiente (mediciones en partes por millón, ppm). Ajuste el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$.

Tóxico	Flujo de la LC50 (y)	LC50 estática (x)
1	23.00	39.00
2	22.30	37.50
3	9.40	22.20
4	9.70	17.50
5	.15	.64
6	.28	.45
7	.75	2.62
8	.51	2.36
9	28.00	32.00
10	.39	.77

a ¿Qué interpretación se puede dar a los resultados?

b Estime el valor de circulación para un tóxico con un valor de LC50 estático de $x = 12$ ppm.

11.9 En la siguiente tabla aparece información acerca de ocho automóviles de cuatro cilindros considerados entre los más eficientes en consumo de combustible en 2006. Los tamaños de los motores se dan en volumen total de cilindros, medido en litros (L).

Automóvil	Volumen de cilindros (x)	Caballos de potencia (y)
Honda Civic	1.8	51
Toyota Prius	1.5	51
VW Golf	2.0	115
VW Beetle	2.5	150
Toyota Corolla	1.8	126
VW Jetta	2.5	150
Mini Cooper	1.6	118
Toyota Yaris	1.5	106

a Localice los puntos en papel milimétrico.

b Encuentre la recta de mínimos cuadrados para los datos.

c Grafique la recta de mínimos cuadrados para ver lo bien que se ajusta a los datos.

d Use la recta de mínimos cuadrados para estimar la clasificación media de potencia para un automóvil eficiente en uso de combustible con volumen de cilindros de 1.9 L.

11.10 Suponga que hemos postulado el modelo

$$Y_i = \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n,$$

donde las ε_i son variables aleatorias independientes y distribuidas idénticamente con $E(\varepsilon_i) = 0$. Entonces $\hat{y}_i = \hat{\beta}_1 x_i$ es el valor pronosticado de y cuando $x = x_i$ y $\text{SSE} = \sum_{i=1}^{n} [y_i - \hat{\beta}_1 x_i]^2$. Encuentre el estimador de mínimos cuadrados de β_1. (Observe que la ecuación $y = \beta x$ describe una recta que pasa por el origen. A menudo el modelo que acabamos de describir se denomina modelo *sin puntos de cruce*.)

11.11 Algunos datos obtenidos por C. E. Marcellari[2] sobre la altura x y diámetro y de caparazones de tortuga aparecen en la tabla siguiente. Si consideramos el modelo

$$E(Y) = \beta_1 x,$$

entonces la pendiente β_1 es la razón entre el diámetro medio y la altura. Use el resultado del Ejercicio 11.10 y los datos siguientes para obtener la estimación de mínimos cuadrados de la relación entre el diámetro medio y la altura.

Espécimen	Diámetro (y)	Altura (x)
OSU 36651	185	78
OSU 36652	194	65
OSU 36653	173	77
OSU 36654	200	76
OSU 36655	179	72
OSU 36656	213	76
OSU 36657	134	75
OSU 36658	191	77
OSU 36659	177	69
OSU 36660	199	65

2. *Fuente*: Carlos E. Marcellari, “Revision of Serpulids of the Genus Rotularia (Annelida) at Seymour Island (Antarctic Peninsula) and Their Value in Stratigraphy,” *Journal of Paleontology* 58(4) (1984).

11.12 Por lo general las procesadoras de alimentos preservan los pepinos fermentándolos en una salmuera baja en sales (6% a 9% de cloruro de sodio) y luego almacenándolos en una salmuera de alto contenido de sales hasta que son utilizados para producir varios tipos de pepinillos en vinagre. La salmuera alta en sales es necesaria para retardar el suavizamiento de los pepinillos y evitar que se congelen cuando se almacenan en el exterior en climas del norte. Los datos que muestran la reducción de la consistencia de los pepinillos almacenados en una salmuera baja en sales (2% a 3%) se dan en la tabla siguiente.[3]

	Semanas (x) en almacenamiento a 72°F				
	0	4	14	32	52
Firmeza (y) en libras	19.8	16.5	12.8	8.1	7.5

a Ajuste una recta de mínimos cuadrados a los datos.

b Para verificar sus cálculos, grafique los cinco puntos que representan a los datos y trace la recta. ¿Le parece que la recta da un buen ajuste de los puntos?

c Use la recta de mínimos cuadrados para estimar la consistencia media de los pepinillos almacenados durante 20 semanas.

11.13 La tabla siguiente proporciona datos sobre la pesca de anchoas en Perú (en millones de toneladas métricas) y los precios de harina de pescado (en dólares actuales por tonelada) durante 14 años consecutivos.[4]

Precio harina de pescado (y)	190	160	134	129	172	197	167
Pesca de anchoas (x)	7.23	8.53	9.82	10.26	8.96	12.27	10.28
Precio harina de pescado (y)	239	542	372	245	376	454	410
Pesca de anchoas (x)	4.45	1.78	4.0	3.3	4.3	0.8	0.5

a Encuentre la recta apropiada de mínimos cuadrados para estos datos.

b Localice los puntos y grafique la recta como comprobación de sus cálculos.

11.14 J. H. Matis y T. E. Wehrly[5] publican la siguiente tabla de datos sobre la proporción de peces de agua dulce que resisten un nivel fijo de contaminación térmica durante lapsos variables.

Proporción de sobrevivientes (y)	Tiempo a escala (x)
1.00	.10
.95	.15
.95	.20
.90	.25
.85	.30
.70	.35
.65	.40
.60	.45
.55	.50
.40	.55

a Ajuste el modelo lineal $Y = \beta_0 + \beta_1 x + \varepsilon$. Dé su interpretación.

b Localice los puntos y grafique el resultado del inciso a. ¿La recta pasa por los puntos?

3. *Fuente*: R. W. Buescher, J. M. Hudson, J. R. Adams, and D. H. Wallace, "Calcium Makes It Possible to Store Cucumber Pickles in Low-Salt Brine", *Arkansas Farm Research* 30(4) (1981).

4. *Fuente*: John E. Bardach and Regina M. Santerre, "Climate and the Fish in the Sea", *BioScience* 31(3) (marzo de 1981): 206ff. Copyright ©1981 by American Institute of Biological Sciences.

5. *Fuente*: J. H. Matis and T. E. Wehrly, "Stochastic Models of Compartmental Systems", *Biometrics* 35(1) (1979): 199-220.

11.4 Propiedades de los estimadores de mínimos cuadrados: regresión lineal simple

Necesitamos determinar las propiedades estadísticas de estimadores de mínimos cuadrados si deseamos usarlos para hacer inferencias estadísticas. En esta sección mostramos que los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$ para los parámetros del modelo lineal simple

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

son estimadores *insesgados* de sus respectivos valores paramétricos. También deduciremos las varianzas de estos estimadores y, dada la suposición de que el término de error ε está distribuido normalmente, demostraremos que $\hat{\beta}_0$ y $\hat{\beta}_1$ tienen distribuciones de muestreo normales. Los resultados correspondientes aplicables al modelo de regresión lineal múltiple se presentan sin demostración en la Sección 11.11.

Recuerde que previamente supusimos que ε es una variable aleatoria con $E(\varepsilon) = 0$. Ahora agregaremos la suposición de que $V(\varepsilon) = \sigma^2$. Esto es, estamos suponiendo que la diferencia entre la variable aleatoria Y y $E(Y) = \beta_0 + \beta_1 x$ está distribuida alrededor de cero con una varianza que no depende de x. Observe que $V(Y) = V(\varepsilon) = \sigma^2$ porque los otros términos del modelo lineal son constantes. (Un estimador insesgado para la varianza σ^2 del término de error del modelo también se presenta en esta sección.)

Suponga que se hacen n observaciones independientes de este modelo para que, antes de muestrear, tengamos n variables aleatorias independientes de la forma

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

De la Sección 11.3 sabemos que

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

que se puede escribir como

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})Y_i - \overline{Y}\sum_{i=1}^{n}(x_i - \overline{x})}{S_{xx}}.$$

Entonces, como $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, tenemos

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})Y_i}{S_{xx}}.$$

Debido a que todas las sumatorias de nuestro análisis se sumarán a partir de $i = 1$ hasta n, simplificaremos nuestra notación al omitir la variable de sumatoria y su índice. Ahora encontremos el valor esperado y la varianza de $\hat{\beta}_1$.

De los teoremas de valores esperados desarrollados en la Sección 5.8 tenemos

$$\begin{aligned} E(\hat{\beta}_1) &= E\left[\frac{\sum(x_i - \overline{x})Y_i}{S_{xx}}\right] = \frac{\sum(x_i - \overline{x})E(Y_i)}{S_{xx}} \\ &= \frac{\sum(x_i - \overline{x})(\beta_0 + \beta_1 x_i)}{S_{xx}} \\ &= \beta_0 \frac{\sum(x_i - \overline{x})}{S_{xx}} + \beta_1 \frac{\sum(x_i - \overline{x})x_i}{S_{xx}}. \end{aligned}$$

Como $\sum(x_i - \overline{x}) = 0$ y $S_{xx} = \sum(x_i - \overline{x})^2 = \sum(x_i - \overline{x})x_i$, tenemos

$$E(\hat{\beta}_1) = 0 + \beta_1 \frac{S_{xx}}{S_{xx}} = \beta_1.$$

Entonces, $\hat{\beta}_1$ es un estimador insesgado de β_1.

Para hallar $V(\hat{\beta}_1)$, usamos el Teorema 5.12. Recuerde que $Y_1, Y_2, \ldots, Y_n$ son independientes y, por tanto,

$$\begin{aligned} V(\hat{\beta}_1) &= V\left[\frac{\sum(x_i - \overline{x})Y_i}{S_{xx}}\right] = \left[\frac{1}{S_{xx}}\right]^2 \sum V[(x_i - \overline{x})Y_i] \\ &= \left[\frac{1}{S_{xx}}\right]^2 \sum (x_i - \overline{x})^2 V(Y_i). \end{aligned}$$

Debido a que $V(Y_i) = \sigma^2$, para $i = 1, 2, \ldots, n$,

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}.$$

Ahora obtendremos el valor esperado y la varianza de $\hat{\beta}_0$, donde $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x}$. Del Teorema 5.12 tenemos

$$V(\hat{\beta}_0) = V(\overline{Y}) + \overline{x}^2 V(\hat{\beta}_1) - 2\overline{x}\,\mathrm{Cov}(\overline{Y}, \hat{\beta}_1).$$

En consecuencia, debemos hallar $V(\overline{Y})$ y $\mathrm{Cov}(\overline{Y}, \hat{\beta}_1)$ para obtener $V(\hat{\beta}_0)$. Como $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, vemos que

$$\overline{Y} = \frac{1}{n}\sum Y_i = \beta_0 + \beta_1\overline{x} + \overline{\varepsilon}.$$

Entonces,

$$E(\overline{Y}) = \beta_0 + \beta_1\overline{x} + E(\overline{\varepsilon}) = \beta_0 + \beta_1\overline{x},$$

y

$$V(\overline{Y}) = V(\overline{\varepsilon}) = \left(\frac{1}{n}\right)V(\varepsilon_1) = \frac{\sigma^2}{n}.$$

Para hallar $\mathrm{Cov}(\overline{Y}, \hat{\beta}_1)$, reescribimos la expresión para $\hat{\beta}_1$ como

$$\hat{\beta}_1 = \sum c_i Y_i,$$

donde

$$c_i = \frac{x_i - \overline{x}}{S_{xx}}.$$

(Observe que $\sum c_i = 0$.) Entonces,

$$\text{Cov}(\overline{Y},\ \hat{\beta}_1) = \text{Cov}\left[\sum\left(\frac{1}{n}\right)Y_i,\ \sum c_i Y_i\right],$$

y usando el Teorema 5.12,

$$\text{Cov}(\overline{Y}, \hat{\beta}_1) = \sum\left(\frac{c_i}{n}\right)V(Y_i) + \sum\sum_{i\neq j}\left(\frac{c_j}{n}\right)\text{Cov}(Y_i, Y_j).$$

Como Y_i y Y_j, donde $i \neq j$, son independientes, Cov $(Y_i, Y_j) = 0$. También, $V(Y_i) = \sigma^2$ y, por tanto,

$$\text{Cov}\,(\overline{Y},\ \hat{\beta}_1) = \frac{\sigma^2}{n}\sum c_i = \frac{\sigma^2}{n}\sum\left(\frac{x_i - \overline{x}}{S_{xx}}\right) = 0.$$

Regresando a nuestra tarea original de hallar el valor esperado y la varianza de $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x}$, aplicamos los teoremas de expectación para obtener

$$E(\hat{\beta}_0) = E(\overline{Y}) - E(\hat{\beta}_1)\overline{x} = \beta_0 + \beta_1\overline{x} - \beta_1\overline{x} = \beta_0.$$

Por tanto, hemos demostrado que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados de sus parámetros respectivos.

Como hemos obtenido $V(\overline{Y})$, $V(\hat{\beta}_1)$ y $\text{Cov}(\overline{Y}, \hat{\beta}_1)$, estamos listos para hallar $V(\hat{\beta}_0)$. Como ya se ha establecido previamente usando el Teorema 5.12,

$$V(\hat{\beta}_0) = V(\overline{Y}) + \overline{x}^2 V(\hat{\beta}_1) - 2\overline{x}\text{Cov}(\overline{Y}, \hat{\beta}_1).$$

Sustituyendo los valores para $V(Y)$, $V(\hat{\beta}_1)$ y $\text{Cov}(\overline{Y}, \hat{\beta}_1)$, obtenemos

$$\begin{aligned} V(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \overline{x}^2\left(\frac{\sigma^2}{S_{xx}}\right) - 0 \\ &= \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right) = \frac{\sigma^2\sum x_i^2}{nS_{xx}}. \end{aligned}$$

Además (vea el Ejercicio 11.21), el Teorema 5.12 puede ser empleado para demostrar que

$$\text{Cov}\left(\hat{\beta}_0,\ \hat{\beta}_1\right) = \frac{-\overline{x}\sigma^2}{S_{xx}}.$$

Observe que $\hat{\beta}_0$ y $\hat{\beta}_1$ están correlacionados (y por tanto son dependientes) a menos que $\overline{x} = 0$.

Todas las cantidades necesarias para determinar los valores de las varianzas y covarianzas anteriores ya han sido calculadas en el curso de obtener los valores para $\hat{\beta}_0$ y $\hat{\beta}_1$.

EJEMPLO **11.2** Encuentre las varianzas de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ para el Ejemplo 11.1.

Solución En el Ejemplo 11.1 (vea los cálculos para el denominador de $\hat{\beta}_1$), encontramos que

$$n = 5, \qquad \sum x_i = 0, \qquad \sum x_i^2 = 10, \qquad S_{xx} = 10.$$

Se deduce que $\overline{x} = 0$,

$$V(\hat{\beta}_0) = \frac{\sigma^2\sum x_i^2}{nS_{xx}} = \frac{\sigma^2(10)}{5(10)} = \left(\frac{1}{5}\right)\sigma^2,$$

y

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} = \left(\frac{1}{10}\right)\sigma^2.$$

Observe que $\text{Cov}(\hat{\beta}_0,\ \hat{\beta}_1) = 0$ en este caso porque $\sum x_i = 0$. ■

Las expresiones anteriores dan las varianzas para los estimadores de mínimos cuadrados en términos de σ^2, la varianza del término de error ε. Por lo general el valor de σ^2 es desconocido y necesitaremos hacer uso de observaciones muestrales para estimar σ^2. Si se usa $\overline{Y}$ para estimar la media, previamente usamos

$$\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

para estimar la varianza poblacional σ^2. Debido a que ahora estamos usando $\hat{Y}_i$ para calcular $E(Y_i)$, parece natural basar una estimación de σ^2 en $\text{SSE} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$. De hecho, demostraremos que

$$S^2 = \left(\frac{1}{n-2}\right)\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \left(\frac{1}{n-2}\right)\text{SSE}$$

proporciona un estimador insesgado para σ^2. Observe que el 2 que se presenta en el denominador de S^2 corresponde al número de parámetros β calculados en el modelo.

Como

$$E(S^2) = E\left[\left(\frac{1}{n-2}\right)\text{SSE}\right] = \left(\frac{1}{n-2}\right)E(\text{SSE}),$$

es necesario hallar $E(\text{SSE})$ para verificar que $E(S^2) = \sigma^2$.

Observe que

$$\begin{aligned}
E(\text{SSE}) &= E\left[\sum(Y_i - \hat{Y}_i)^2\right] = E\left[\sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right] \\
&= E\left[\sum(Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2\right] \\
&= E\left[\sum[(Y_i - \overline{Y}) - \hat{\beta}_1(x_i - \overline{x})]^2\right] \\
&= E\left[\sum(Y_i - \overline{Y})^2 + \hat{\beta}_1^2\sum(x_i - \overline{x})^2 - 2\hat{\beta}_1\sum(x_i - \overline{x})(Y_i - \overline{Y})\right].
\end{aligned}$$

Como $\sum(x_i - \overline{x})(Y_i - \overline{Y}) = \sum(x_i - \overline{x})^2\hat{\beta}_1$, los últimos dos términos en la expectación se combinan para dar $-\hat{\beta}_1^2\sum(x_i - \overline{x})^2$. También,

$$\sum(Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2,$$

y, por tanto,

$$\begin{aligned}
E\left[\sum(Y_i - \hat{Y}_i)^2\right] &= E\left[\sum Y_i^2 - n\overline{Y}^2 - \hat{\beta}_1^2 S_{xx}\right] \\
&= \sum E\left(Y_i^2\right) - nE\left(\overline{Y}^2\right) - S_{xx}E\left(\hat{\beta}_1^2\right).
\end{aligned}$$

Si observamos que, para cualquier variable aleatoria U, $E(U^2) = V(U) + [E(U)]^2$, vemos que

$$
\begin{aligned}
E\left[\sum(Y_i - \hat{Y}_i)^2\right] &= \sum\{V(Y_i) + [E(Y_i)]^2\} - n\{V(\overline{Y}) + [E(\overline{Y})]^2\} \\
&\quad - S_{xx}\{V(\hat{\beta}_1) + [E(\hat{\beta}_1)]^2\} \\
&= n\sigma^2 + \sum(\beta_0 + \beta_1 x_i)^2 - n\left[\frac{\sigma^2}{n} + (\beta_0 + \beta_1\overline{x})^2\right] \\
&\quad - S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta_1^2\right).
\end{aligned}
$$

Esta expresión se simplifica a $(n-2)\sigma^2$. Entonces, encontramos que un estimador insesgado de σ^2 está dado por

$$S^2 = \left(\frac{1}{n-2}\right)\sum(Y_i - \hat{Y}_i)^2 = \left(\frac{1}{n-2}\right)\text{SSE}.$$

Queda pendiente una tarea, hallar una forma fácil de calcular $\sum(y_i - \hat{y}_i)^2 = \text{SSE}$. En el Ejercicio 11.15(a) se demostrará que una fórmula de cálculo para la SSE está dada por

$$
\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n}(y_i - \overline{y})^2 - \hat{\beta}_1\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \\
&= S_{yy} - \hat{\beta}_1 S_{xy}, \qquad \text{donde } S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2.
\end{aligned}
$$

EJEMPLO **11.3** Calcule σ^2 con los datos dados en el Ejemplo 11.1.

Solución Para estos datos, $n = 5$ y ya hemos determinado que

$$\sum y_i = 5, \qquad S_{xy} = 7, \qquad \hat{\beta}_1 = .7.$$

Fácilmente se determina que $\sum y_i^2 = 11$ y que

$$S_{yy} = \sum(y_i - \overline{y})^2 = \sum y_i^2 - n(\overline{y})^2 = 11 - 5(1)^2 = 6.0.$$

Por tanto,

$$\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy} = 6.0 - (.7)(7) = 1.1,$$

y

$$s^2 = \frac{\text{SSE}}{n-2} = \frac{1.1}{5-2} = \frac{1.1}{3} = .367.$$

■

Estas deducciones establecen las medias y las varianzas de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ y demuestran que $S^2 = \text{SSE}/(n-2)$ es un estimador insesgado para el parámetro σ^2. Hasta este punto, las únicas suposiciones que hemos hecho acerca del término de error ε del modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ son que $E(\varepsilon) = 0$ y que $V(\varepsilon) = \sigma^2$, independiente de x. La forma de las distribuciones muestrales para $\hat{\beta}_0$ y $\hat{\beta}_1$ depende de la distribución del término de error ε. Debido a que con frecuencia se presenta la distribución normal en la naturaleza, es razonable suponer que ε está distribuido normalmente con media 0 y varianza σ^2.

Si se garantiza esta suposición de normalidad, se deduce que Y_i está distribuida normalmente con media $\beta_0 + \beta_1 x_i$ y varianza σ^2. Debido a que $\hat{\beta}_0$ y $\hat{\beta}_1$ son *funciones lineales* de $Y_1, Y_2, \ldots, Y_n$, los estimadores están distribuidos normalmente, con las medias y varianzas que se determinaron previamente. Además, si se garantiza la suposición de normalidad, se deduce que

$$\frac{(n-2)S^2}{\sigma^2} = \frac{\text{SSE}}{\sigma^2}$$

tiene una distribución χ^2 con $n - 2$ grados de libertad. (La prueba de este resultado se omite.)

Como veremos más adelante, la suposición de normalidad de la distribución del término de error ε y las distribuciones normales resultantes para $\hat{\beta}_0$ y $\hat{\beta}_1$ nos permitirán desarrollar pruebas e intervalos de confianza basados en la distribución t. Los resultados de esta sección se resumen aquí por su importancia para las discusiones en secciones subsiguientes. Observe que $V(\hat{\beta}_0)$, $V(\hat{\beta}_1)$ y $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ son todos múltiplos constantes de σ^2. Como $V(\hat{\beta}_i) = \text{Cov}(\hat{\beta}_i, \hat{\beta}_i)$, unificaremos la notación y daremos coherencia a las últimas secciones de este capítulo si usamos la notación $V(\hat{\beta}_0) = c_{00}\sigma^2$, $V(\hat{\beta}_1) = c_{11}\sigma^2$ y $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$.

Propiedades de los estimadores de mínimos cuadrados; regresión lineal simple

1. Los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son insesgados, es decir, $E(\hat{\beta}_i) = \beta_i$, para $i = 0, 1$.
2. $V(\hat{\beta}_0) = c_{00}\sigma^2$, donde $c_{00} = \sum x_i^2/(nS_{xx})$.
3. $V(\hat{\beta}_1) = c_{11}\sigma^2$, donde $c_{11} = \dfrac{1}{S_{xx}}$.
4. $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$, donde $c_{01} = \dfrac{-\overline{x}}{S_{xx}}$.
5. Un estimador insesgado de σ^2 es $S^2 = \text{SSE}/(n-2)$, donde $\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy}$ y $S_{yy} = \sum (y_i - \overline{y})^2$.

Si, además, el ε_i, para $i = 1, 2, \ldots, n$ está distribuido normalmente,

6. $\hat{\beta}_0$ y $\hat{\beta}_1$ están distribuidas normalmente.
7. La variable aleatoria $\dfrac{(n-2)S^2}{\sigma^2}$ tiene una distribución χ^2 con $n - 2$ grados de libertad.
8. El estadístico S^2 es independiente de $\hat{\beta}_0$ y $\hat{\beta}_1$.

Ejercicios

11.15 **a** Deduzca la siguiente identidad:

$$\begin{aligned}\text{SSE} &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= \sum_{i=1}^{n}(y_i - \overline{y})^2 - \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = S_{yy} - \hat{\beta}_1 S_{xy}.\end{aligned}$$

Observe que esto proporciona un método computacional más fácil para hallar la SSE.

b Use la fórmula computacional para la suma de cuadrados del error (SSE), deducida en el inciso a, para demostrar que $\text{SSE} \le S_{yy}$. [*Sugerencia:* $\hat{\beta}_1 = S_{xy}/S_{xx}$.]

11.16 Se realizó un experimento para observar el efecto de un aumento en temperatura en la potencia de un antibiótico. Tres porciones de 1 onza del antibiótico se almacenaron durante tiempos iguales a cada una de las siguientes temperaturas Fahrenheit: 30°, 50°, 70° y 90°. Las lecturas de potencia observadas al final del periodo experimental fueron como se muestra en la tabla siguiente.

Lecturas de potencia (y)	38, 43, 29	32, 26, 33	19, 27, 23	14, 19, 21
Temperatura (x)	30°	50°	70°	90°

a Encuentre la recta de mínimos cuadrados apropiada para estos datos.

b Localice los puntos y grafique la recta como comprobación de los cálculos.

c Calcule S^2.

11.17 **a** Calcule la SSE y S^2 para el Ejercicio 11.5.

b Algunas veces es conveniente para fines de cálculo, tener valores x espaciados simétrica e igualmente alrededor de cero. Los valores x se pueden cambiar de escala (o codificarse) en cualquier forma conveniente sin pérdida de información en el análisis estadístico. Consulte el Ejercicio 11.5. Codifique los valores x (originalmente dados en una escala de 1 a 8) con el uso de la fórmula

$$x^* = \frac{x - 4.5}{.5}.$$

A continuación ajuste el modelo $Y = \beta_0^* + \beta_1^* x^* + \varepsilon$. Calcule la SSE. (Observe que los valores x^* son enteros simétricamente espaciados alrededor de cero.) Compare la SSE con el valor obtenido en el inciso a.

11.18 **a** Calcule la SSE y S^2 para el Ejercicio 11.8.

b Consulte el Ejercicio 11.8. Codifique los valores x en una forma conveniente y ajuste un modelo lineal simple a las mediciones de LC50 presentadas ahí. Calcule la SSE y compare su respuesta con el resultado del inciso a.

11.19 Se realizó un estudio para determinar los efectos de la privación de sueño en la capacidad de las personas para resolver problemas sencillos. La cantidad de privación de sueño varió en 8, 12, 16, 20 y 24 horas sin dormir. Un total de diez individuos participaron en el estudio, dos por cada nivel de privación de sueño. Después de su periodo de privación de sueño, a cada individuo se le presentó un conjunto de problemas sencillos de sumas para que lo resolvieran, registrándose el número de errores; se obtuvieron los resultados mostrados en la tabla siguiente.

Número de errores (y)	8, 6	6, 10	8, 14	14, 12	16, 12
Número de horas sin dormir (x)	8	12	16	20	24

a Encuentre la recta de mínimos cuadrados apropiada para estos datos.

b Localice los puntos y grafique la recta de mínimos cuadrados en sus cálculos.

c Calcule S^2.

11.20 Suponga que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias normales e independientes con $E(Y_i) = \beta_0 + \beta_1 x_i$ y $V(Y_i) = \sigma^2$, para $i = 1, 2, \ldots, n$. Demuestre que los estimadores de máxima probabilidad (MLE) de β_0 y β_1 son iguales que los estimadores de mínimos cuadrados de la Sección 11.3.

11.21 De acuerdo con las suposiciones del Ejercicio 11.20, encuentre $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$. Use esta respuesta para demostrar que $\hat{\beta}_0$ y $\hat{\beta}_1$ son independientes si $\sum_{i=1}^n x_i = 0$. [*Sugerencia:* $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \mathrm{Cov}(\overline{Y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1)$. Use el Teorema 5.12 y los resultados de esta sección.]

11.22 De acuerdo con las suposiciones del Ejercicio 11.20, encuentre el MLE de σ^2.

11.5 Inferencias respecto a los parámetros β_i

Suponga que un ingeniero ha ajustado el modelo

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

donde Y es la resistencia del concreto después de 28 días y x es la razón entre agua y cemento empleada en el concreto. Si, en realidad, la resistencia del concreto no cambia con la razón entre agua y cemento, entonces $\beta_1 = 0$. Así, el ingeniero puede probar H_0: $\beta_1 = 0$ contra H_a: $\beta_1 \neq 0$ para evaluar si la variable independiente tiene una influencia en la variable dependiente. O bien, el ingeniero puede estimar la tasa media de cambio β_1 en $E(Y)$ para un cambio de 1 unidad en la proporción x entre agua y cemento.

En general, para cualquier modelo de regresión lineal, si el error aleatorio ε está distribuido normalmente, hemos establecido que $\hat{\beta}_i$ es un estimador de β_i insesgado y distribuido normalmente con

$$V(\hat{\beta}_0) = c_{00}\sigma^2, \qquad \text{donde } c_{00} = \frac{\sum x_i^2}{nS_{xx}}$$

y

$$V(\hat{\beta}_1) = c_{11}\sigma^2, \qquad \text{donde } c_{11} = \frac{1}{S_{xx}}.$$

Esto es, las varianzas de ambos estimadores son múltiplos constantes de σ^2, la varianza del término de error del modelo. Usando esta información, podemos construir una prueba de la hipótesis H_0: $\beta_i = \beta_{i0}$ (β_{i0} es un valor específico de β_i), usando el estadístico de prueba

$$Z = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}},$$

donde

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \quad \text{y} \quad c_{11} = \frac{1}{S_{xx}}.$$

La región de rechazo para una prueba de dos colas está dada por

$$|z| \geq z_{\alpha/2}.$$

Como en el caso de las pruebas Z simples estudiadas en el Capítulo 10, para calcular cualquiera de los estadísticos Z precedentes debemos conocer σ o tener una buena estimación basada en un número adecuado de grados de libertad. (Lo que sería adecuado es un punto discutible. Sugerimos que la estimación sea basada en 30 o más grados de libertad.) Cuando esta estimación no exista (que por lo general es el caso), puede calcularse una estimación de σ a partir de los datos experimentales (de acuerdo con el procedimiento de la Sección 11.4) y sustituirse por σ en el estadístico Z. Si estimamos σ con $S = \sqrt{\text{SSE}/(n-2)}$, puede demostrarse que la cantidad resultante

$$T = \frac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$$

tiene una distribución t de Student con $n - 2$ grados de libertad (véase el Ejercicio 11.27).

Prueba de hipótesis para β_i

$H_0 : \beta_i = \beta_{i0}$.

$$H_a : \begin{cases} \beta_i > \beta_{i0} & \text{(región de rechazo de cola superior),} \\ \beta_i < \beta_{i0} & \text{(región de rechazo de cola inferior),} \\ \beta_i \neq \beta_{i0} & \text{(región de rechazo de dos colas).} \end{cases}$$

Estadístico de prueba: $T = \dfrac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$.

$$\text{Región de rechazo:} \begin{cases} t > t_\alpha & \text{(alternativa de cola superior),} \\ t < -t_\alpha & \text{(alternativa de cola inferior),} \\ |t| > t_{\alpha/2} & \text{(alternativa de dos colas),} \end{cases}$$

donde

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \quad \text{y} \quad c_{11} = \frac{1}{S_{xx}}.$$

Observe que t_α está basada en $(n - 2)$ grados de libertad.

EJEMPLO **11.4** ¿Los datos del Ejemplo 11.1 presentan suficiente evidencia para indicar que la pendiente difiere de 0? Pruebe usando $\alpha = .05$ y establezca los límites para el nivel de significancia alcanzado.

Solución La pregunta anterior supone que el modelo probabilístico es una descripción realista de la verdadera respuesta e implica una prueba de hipótesis $H_0: \beta_1 = 0$ contra $H_a: \beta_1 \neq 0$ del modelo lineal $Y = \beta_0 + \beta_1 x + \varepsilon$. Para estos datos, determinamos en el Ejemplo 11.1 que $\hat{\beta}_1 = .7$ y $S_{xx} = 10$. El Ejemplo 11.3 dio como resultado $s^2 = \text{SSE}/(n-2) = .367$ y $s = \sqrt{.367} = .606$. (*Nota*: la SSE está basada en $n - 2 = 3$ grados de libertad).

Como estamos interesados en el parámetro β_1, necesitamos el valor

$$c_{11} = \frac{1}{S_{xx}} = \frac{1}{10} = .1.$$

Entonces,

$$t = \frac{\hat{\beta}_1 - 0}{s\sqrt{c_{11}}} = \frac{.7 - 0}{.606\sqrt{.1}} = 3.65.$$

Si tomamos $\alpha = .05$, el valor de $t_{\alpha/2} = t_{.025}$ para 3 grados de libertad es 3.182, y la región de rechazo es

$$\text{rechazar si } |t| \geq 3.182.$$

Debido a que el valor absoluto del valor calculado de t es mayor que 3.182, rechazamos la hipótesis nula de que $\beta_1 = 0$ en el nivel de significancia $\alpha = .05$. Como la prueba es de dos colas, el valor $p = 2P(t > 3.65)$, donde t tiene una distribución t con 3 grados de libertad. Usando la Tabla 5, Apéndice 3, encontramos que $.01 < P(t > 3.65) < .025$. Entonces, concluimos que $.02 <$ valor $p < .05$. En consecuencia, rechazaríamos la hipótesis nula para cualquier valor de

$\alpha \geq .05$. Para valores de $\alpha \leq .02$, no rechazaríamos la hipótesis nula. Si hubiéramos escogido $.02 < \alpha < .05$, se requeriría información más específica acerca del valor p. La aplicación *Student's t Probabilities and Quantiles* dice que, con 3 grados de libertad, el valor $p = 2P(t > 3.65) = 2(.01775) = .0355$. De nuevo observamos el acuerdo entre las conclusiones alcanzadas por el procedimiento de prueba formal (α fija) y la interpretación correcta del nivel de significancia alcanzado.

Como paso adicional del análisis, podríamos ver el ancho de un intervalo de confianza para β_1 para ver si es lo suficientemente corto para detectar una desviación de cero que sería de significancia práctica. Demostraremos que el intervalo de confianza para β_1 es bastante ancho, lo que sugiere que el experimentador necesita recolectar más información antes de tomar una decisión. ■

Con base en el estadístico t dado antes, podemos seguir los procedimientos del Capítulo 10 para demostrar que un intervalo de confianza para β_i, con coeficiente de confianza $1 - \alpha$, está dado por:

Un intervalo de confianza 100(1 – α)% para β_i

$$\hat{\beta}_i \pm t_{\alpha/2} S\sqrt{c_{ii}},$$

donde

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \quad \text{y} \quad c_{11} = \frac{1}{S_{xx}}.$$

EJEMPLO **11.5** Calcule un intervalo de confianza de 95% para el parámetro β_1 del Ejemplo 11.4.

Solución El valor tabulado para $t_{.025}$, basado en 3 grados de libertad, es 3.182. Entonces el intervalo de confianza de 95% para β_1 es

$$\hat{\beta}_1 \pm t_{.025} s\sqrt{c_{11}}.$$

Sustituyendo, obtenemos

$$.7 \pm (3.182)(.606)\sqrt{0.1}, \quad \text{o} \quad .7 \pm .610.$$

Si deseamos estimar β_1 correcta con tolerancia de no más de .15 unidad, es obvio que el intervalo de confianza es demasiado ancho y que el tamaño muestral debe ser aumentado. ■

Ejercicios

11.23 Consulte el Ejercicio 11.3

a ¿Los datos presentan suficiente evidencia para indicar que la pendiente β_1 difiere de cero? (Pruebe en el nivel de significancia de 5%.)

b ¿Qué se puede decir acerca del nivel de significancia alcanzado asociado con la prueba implantada en el inciso a usando una tabla del apéndice?

c **Ejercicio Applet** ¿Qué se puede decir acerca del nivel de significancia alcanzado asociado con la prueba realizada en el inciso a usando la aplicación apropiada?

d Encuentre un intervalo de confianza de 95% para β_1.

11.24 Consulte el Ejercicio 11.13. ¿Los datos presentan suficiente evidencia para indicar que el tamaño x de la pesca de anchoas aporta información para la predicción del precio y de la harina de pescado?

a Determine los límites sobre el nivel de significancia alcanzado.

b **Ejercicio Applet** ¿Cuál es el valor p exacto?

c Con base en sus respuestas a los incisos a y/o b, ¿qué concluiría en el nivel de significancia de $\alpha = .10$?

11.25 ¿Los datos del Ejercicio 11.19 presentan suficiente evidencia para indicar que el número de errores está relacionado linealmente con el número de horas sin dormir?

a Determine los límites sobre el nivel de significancia alcanzado.

b **Ejercicio Applet** Determine el valor p exacto.

c Con base en sus respuestas a los incisos a y/o b, ¿qué concluiría en el nivel de significancia de $\alpha = .05$?

d ¿Esperaría que la relación entre y y x fuera lineal si x se hace variar en una amplitud más ancha, por ejemplo de $x = 4$ a $x = 48$?

e Obtenga un intervalo de confianza de 95% para la pendiente. Dé una interpretación práctica para esta estimación de intervalo.

11.26 A la mayoría de los estudiantes de segundo año de física se les pide que realicen un experimento para verificar la ley de Hooke. La ley de Hooke establece que cuando se aplica una fuerza a un cuerpo que es largo en comparación con su área de sección transversal, el cambio y en su longitud es proporcional a la fuerza x; esto es,

$$y = \beta_1 x,$$

donde β_1 es una constante de proporcionalidad. Los resultados del experimento de laboratorio de un estudiante de física se muestran en la siguiente tabla. Se usaron seis tramos de alambre de acero, de .34 milímetro (mm) de diámetro y 2 metros (m) de largo, para obtener las seis mediciones de cambio de fuerza-longitud.

Fuerza x (kg)	Cambio en longitud (y) (mm)
29.4	4.25
39.2	5.25
49.0	6.50
58.8	7.85
68.6	8.75
78.4	10.00

a Ajuste el modelo, $Y = \beta_0 + \beta_1 x + \varepsilon$, a los datos, usando el método de mínimos cuadrados.

b Encuentre un intervalo de confianza de 95% para la pendiente de la recta.

c De acuerdo con la ley de Hooke, la recta debe pasar por el punto (0, 0); esto es, β_0 debe ser igual a 0. Pruebe la hipótesis de que $E(Y) = 0$ cuando $x = 0$. Determine los límites para el nivel de significancia alcanzado.

d **Ejercicio Applet** ¿Cuál es el valor p exacto?

e ¿Qué concluiría usted en nivel $\alpha = .05$?

11.27 Use las propiedades de los estimadores de mínimos cuadrados dados en la Sección 11.4 para completar lo siguiente.

a Demuestre que dada la hipótesis nula $H_0: \beta_i = \beta_{i0}$

$$T = \frac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$$

tiene una distribución t con $n - 2$ grados de libertad, donde $i = 1, 2$.

b Deduzca los intervalos de confianza para β_i dados en esta sección.

11.28 Suponga que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias distribuidas normalmente e independientes, con $E(Y_i) = \beta_0 + \beta_1 x_i$ y $V(Y_i) = \sigma^2$, para $i = 1, 2, \ldots, n$. Demuestre que la prueba de razón de verosimilitudes de $H_0: \beta_1 = 0$ contra $H_a: \beta_1 \neq 0$ es equivalente a la prueba t dada en esta sección.

***11.29** Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias como las dadas en el Ejercicio 11.28. Suponga que tenemos un conjunto adicional de variables aleatorias independientes $W_1, W_2, \ldots, W_m$, donde W_i está distribuida normalmente con $E(W_i) = \gamma_0 + \gamma_1 c_i$ y $V(W_i) = \sigma^2$, para $i = 1, 2, \ldots, m$. Construya una prueba de H_0: $\beta_1 = \gamma_1$ contra $H_a: \beta_1 \neq \gamma_1$.[6]

11.30 El octanaje Y del petróleo refinado está relacionado con la temperatura x del proceso de refinación, pero también está relacionado con el tamaño de partículas del catalizador. Un experimento con un catalizador de partículas pequeñas dio como resultado una recta de mínimos cuadrados de

$$\hat{y} = 9.360 + .155x,$$

con $n = 31$, $V(\hat{\beta}_1) = (.0202)^2$ y SSE $= 2.04$. Un experimento independiente con un catalizador de partículas grandes dio como resultado

$$\hat{y} = 4.265 + .190x,$$

con $n = 11$, $V(\hat{\beta}_1) = (.0193)^2$ y SSE $= 1.86$.[7]

a Pruebe las hipótesis de que las pendientes son considerablemente diferentes de cero, con cada prueba en el nivel de significancia de .05.

***b** Pruebe con un nivel de significancia de .05 que los dos tipos de catalizador producen la misma pendiente en la relación entre el octanaje y la temperatura. (Use la prueba que desarrolló en el Ejercicio 11.29.)

11.31 Usando un procedimiento químico llamado *polarografía de pulso diferencial*, un químico midió la corriente pico generada (en microamperes, μA) cuando soluciones que contienen diferentes cantidades de níquel (medidas en partes por mil millones, ppmm) se agregan a diferentes porciones de la misma solución amortiguadora.[8] ¿Hay suficiente evidencia para indicar que la corriente pico aumenta cuando aumentan las concentraciones de níquel? Use $\alpha = .05$.

x = Ni (ppmm)	y = Corriente pico(μA)
19.1	.095
38.2	.174
57.3	.256
76.2	.348
95	.429
114	.500
131	.580
150	.651
170	.722

6. Los ejercicios precedidos por un asterisco son opcionales.

7. *Fuente*: Gweyson and Cheasley, *Petroleum Refiner* (agosto de 1959): 135.

8. *Fuente*: Daniel C. Harris, *Quantitative Chemical Analysis*, 3rd ed. (New York, Freeman, 1991).

11.32 Consulte los Ejercicios 11.5 y 11.17.

a ¿Hay suficiente evidencia para indicar que la mediana de los precios de venta de casas unifamiliares nuevas aumentó en el periodo de 1972 a 1979 en el nivel de significancia de .01?

b Estime el aumento anual esperado en la mediana de los precios de venta construyendo un intervalo de confianza de 99%.

11.33 Consulte los Ejercicios 11.8 y 11.18. ¿Hay suficiente evidencia de una relación lineal entre las LC50 de flujo y estáticas? Pruebe al nivel de significancia de .05.

11.34 Consulte el Ejercicio 11.33. ¿Hay suficiente evidencia de una relación lineal entre las LC50 de flujo y estática?

a Determine los límites para el nivel de significancia alcanzado.

b **Ejercicio Applet** ¿Cuál es el valor p exacto?

11.6 Inferencias respecto a funciones lineales de los parámetros del modelo: regresión lineal simple

Además de hacer inferencias acerca de una β_i individual, a menudo nos interesa hacer inferencias acerca de funciones lineales de los parámetros del modelo β_0 y β_1. Por ejemplo, podríamos calcular $E(Y)$, dada por

$$E(Y) = \beta_0 + \beta_1 x,$$

donde $E(Y)$ representa el rendimiento medio de un proceso químico para los ajustes de la variable x de un proceso controlado, o el rendimiento de combustible por milla de recorrido de motores de gasolina de cuatro cilindros con volumen x de cilindrada. Las propiedades de los estimadores de estas funciones lineales se establecen en esta sección.

Suponga que deseamos hacer una inferencia acerca de la función lineal

$$\theta = a_0\beta_0 + a_1\beta_1,$$

donde a_0 y a_1 son constantes (una de las cuales puede ser igual a cero). Entonces, la misma función lineal de los estimadores de los parámetros

$$\hat{\theta} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1,$$

es un estimador insesgado de θ ya que, por el Teorema 5.12,

$$E(\hat{\theta}) = a_0 E(\hat{\beta}_0) + a_1 E(\hat{\beta}_1) = a_0\beta_0 + a_1\beta_1 = \theta.$$

Si aplicamos el mismo teorema, determinamos que la varianza de $\hat{\theta}$ es

$$V(\hat{\theta}) = a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) + 2a_0a_1\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1),$$

donde $V(\hat{\beta}_i) = c_{ii}\sigma^2$ y $\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$, con

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}}, \qquad c_{11} = \frac{1}{S_{xx}}, \qquad c_{01} = \frac{-\overline{x}}{S_{xx}}.$$

Con algunas manipulaciones algebraicas de rutina tendremos

$$V(\hat{\theta}) = \left(\frac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}} \right) \sigma^2.$$

Por último, recordando que $\hat{\beta}_0$ y $\hat{\beta}_1$ están distribuidas normalmente en muestreo repetido (Sección 11.4), es evidente que $\hat{\theta}$ es una función lineal de variables aleatorias distribuidas normalmente, lo cual implica que $\hat{\theta}$ está distribuida normalmente.

Entonces, concluimos que

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

tiene una distribución normal y podría emplearse para probar la hipótesis

$$H_0 : \theta = \theta_0$$

cuando θ_0 es algún valor especificado de $\theta = a_0\beta_0 + a_1\beta_1$. Del mismo modo, un intervalo de confianza de $100(1 - \alpha)\%$ para $\theta = a_0\beta_0 + a_1\beta_1$ es

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}.$$

Observamos que tanto en el estadístico Z como en el intervalo de confianza inmediato anterior, $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$ es un múltiplo constante de σ (dependiendo del tamaño muestral n, de los valores de las x y de los valores de las a). Si sustituimos S por σ en la expresión para Z, la expresión resultante (que identificamos como T) tiene una distribución t de Student en muestreo repetido, con $n - 2$ grados de libertad y proporciona un estadístico de prueba para verificar la hipótesis acerca de $\theta = a_0\beta_0 + a_1\beta_1$.

A continuación se resumen pruebas adecuadas.

Una prueba para $\boldsymbol{\theta} = a_0\beta_0 + a_1\beta_1$

$$H_0 : \theta = \theta_0,$$

$$H_a : \begin{cases} \theta > \theta_0, \\ \theta < \theta_0, \\ \theta \neq \theta_0. \end{cases}$$

Estadístico de prueba: $$T = \frac{\hat{\theta} - \theta_0}{S\sqrt{\left(\frac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}} \right)}}.$$

Región de rechazo: $$\begin{cases} t > t_\alpha, \\ t < -t_\alpha, \\ |t| > t_{\alpha/2}. \end{cases}$$

Aquí, t_α y $t_{\alpha/2}$ están basados en $n - 2$ grados de libertad.

El intervalo de confianza correspondiente de 100(1 – α)% para $\theta = a_0\beta_0 + a_1\theta_1$ es el siguiente.

Un intervalo de confianza 100(1 – α)% para $\theta = a_0\beta_0 + a_1\beta_1$

$$\hat{\theta} \pm t_{\alpha/2} S \sqrt{\left(\frac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}}\right)},$$

donde la $t_{\alpha/2}$ tabulada está basada en n – 2 grados de libertad.

Una aplicación útil de las técnicas de prueba de hipótesis e intervalo de confianza que acabamos de presentar es el problema de estimar $E(Y)$, el valor medio de Y, para un valor fijo de la variable independiente x. En particular, si x^* denota un valor específico de x que es de interés, entonces

$$E(Y) = \beta_0 + \beta_1 x^*.$$

Observe que $E(Y)$ es un caso especial de $a_0\beta_0 + a_1\beta_1$, con $a_0 = 1$ y $a_1 = x^*$. Entonces, se puede hacer una inferencia acerca de $E(Y)$ cuando $x = x^*$ usando las técnicas desarrolladas anteriormente para combinaciones lineales generales de las β.

En el contexto de estimar el valor medio de Y, $E(Y) = \beta_0 + \beta_1 x^*$ cuando la variable independiente x toma el valor x^*, se puede demostrar (vea el Ejercicio 11.35) que, con $a_0 = 1$, $a_1 = x^*$,

$$\left(\frac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}}\right) = \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}.$$

Un intervalo de confianza para el valor medio de Y cuando $x = x^*$, un valor particular de x es el siguiente.

Un intervalo de confianza 100(1 – α)% para $E(Y) = \beta_0 + \beta_1 x^*$

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}},$$

donde la $t_{\alpha/2}$ tabulada está basada en n – 2 grados de libertad.

Esta fórmula permite ver con facilidad que para un valor fijo de n y para valores x dados, el intervalo de confianza más corto para $E(Y)$ se obtiene cuando $x^* = \overline{x}$, el promedio de los valores x empleados en el experimento. Si nuestro objetivo es planear un experimento que arroje intervalos de confianza cortos para $E(Y)$ cuando $x = x^*$, n debe ser grande, S_{xx} debe ser grande (si es posible) y $\overline{x}$ debe ser cercana a x^*. La interpretación física de una S_{xx} grande es que, cuando sea posible, los valores de x empleados en el experimento deben estar tan *dispersos* como sea posible.

EJEMPLO **11.6** Para los datos del Ejemplo 11.1, determine un intervalo de confianza de 90% para $E(Y)$ cuando $x = 1$.

Solución Para el modelo del Ejemplo 11.1,

$$E(Y) = \beta_0 + \beta_1 x.$$

Para estimar $E(Y)$ para cualquier valor fijo $x = x^*$, usamos el estimador insesgado $\widehat{E(Y)} = \hat{\beta}_0 + \hat{\beta}_1 x^*$. Entonces,

$$\hat{\beta}_0 + \hat{\beta}_1 x^* = 1 + .7x^*.$$

Para este caso, $x^* = 1$; y como $n = 5$, $\overline{x} = 0$ y $S_{xx} = 10$, se deduce que

$$\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}} = \frac{1}{5} + \frac{(1-0)^2}{10} = .3.$$

En el Ejemplo 11.3 encontramos que s^2 es .367 o $s = .606$, para estos datos. El valor de $t_{.05}$ con $n - 2 = 3$ grados de libertad es 2.353.

El intervalo de confianza para $E(Y)$ cuando $x = 1$ es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

$$[(1 + (.7)(1)] \pm (2.353)(.606)\sqrt{.3}$$

$$1.7 \pm .781.$$

Esto es, tenemos confianza de 90% en que, cuando la variable independiente tome el valor de $x = 1$, el valor medio $E(Y)$ de la variable dependiente es entre .919 y 2.481. Este intervalo obviamente es muy ancho, pero recuerde que está basado en sólo cinco puntos y se empleó sólo para fines de ilustración. Demostraremos algunas aplicaciones prácticas del análisis de regresión en la Sección 11.9. ■

Ejercicios

11.35 Para el modelo de regresión lineal simple $Y = \beta_0 + \beta_1 x + \varepsilon$ con $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma^2$, use la expresión para $V(a_0\hat{\beta}_0 + a_1\hat{\beta}_1)$ deducida en esta sección para demostrar que

$$V(\hat{\beta}_0 + \hat{\beta}_1 x^*) = \left[\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}\right]\sigma^2.$$

¿Para qué valor de x^* alcanza su longitud mínima el intervalo de confianza para $E(Y)$?

11.36 Consulte los Ejercicios 11.13 y 11.24. Encuentre el intervalo de confianza de 90% para el precio medio por tonelada de harina de pescado si la captura de anchoas es de 5 millones de toneladas métricas.

11.37 Usando el modelo ajustado para los datos del Ejercicio 11.8 construya un intervalo de confianza de 95% para el valor medio de LC50 de flujo para un tóxico que tiene un LC50 estático de 12 partes por millón. (Véase también el Ejercicio 11.18.)

11.38 Consulte el Ejercicio 11.3. Encuentre un intervalo de confianza de 90% para $E(Y)$ cuando $x^* = 0$. Entonces encuentre intervalos de confianza de 90% para $E(Y)$ cuando $x^* = -2$ y $x^* = +2$. Compare las amplitudes de estos intervalos. Localice estos límites de confianza en la gráfica construida en el Ejercicio 11.3.

11.39 Consulte el Ejercicio 11.16. Determine un intervalo de confianza de 95% para la potencia media de una onza de antibiótico almacenado a 65°F.

11.40 Consulte el Ejercicio 11.14. Determine un intervalo de confianza de 90% para la proporción esperada de sobrevivientes en el periodo .30.

***11.41** Consulte el Ejercicio 11.4. Suponga que la muestra dada ahí provino de una población grande pero finita de artículos de inventario. Deseamos calcular la media poblacional de los valores auditados usando el hecho de que los valores en libros se conocen para cualquier artículo en inventario. Si la población contiene N artículos y

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_i,$$

entonces la media poblacional está dada por

$$\mu_Y = \frac{1}{N}\sum_{i=1}^{N}\mu_i = \beta_0 + \beta_1\left(\frac{1}{N}\right)\sum_{i=1}^{N}x_i = \beta_0 + \beta_1\mu_x.$$

a Usando los estimadores de mínimos cuadrados de β_0 y β_1, demuestre que μ_Y puede ser estimada por

$$\hat{\mu}_Y = \overline{y} + \hat{\beta}_1(\mu_x - \overline{x}).$$

(Observe que $\overline{y}$ se ajusta arriba o abajo, dependiendo de si $\overline{x}$ es mayor o menor que μ_x.)

b Usando los datos del Ejercicio 11.4 y el hecho de que $\mu_x = 74.0$, calcule μ_Y, la media de los valores auditados, y ponga un límite de desviación estándar de 2 en el error de estimación. (Considere los valores de x_i como constantes cuando calcule la varianza de $\hat{\mu}_Y$.)

11.7 Predicción de un valor particular de Y mediante regresión lineal simple

Suponga que para una presión fija, el rendimiento Y para un experimento químico es una función de la temperatura x a la que el experimento se efectúa. Suponga que un modelo lineal de la forma

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

representa adecuadamente la función de respuesta trazada por Y sobre la región experimental de interés. En la Sección 11.6 estudiamos métodos para estimar el rendimiento medio $E(Y)$ del proceso en el ajuste $x = x^*$.

Ahora considere un problema diferente. En lugar de calcular el rendimiento medio en x^*, deseamos *predecir* la respuesta particular Y que observaremos si el experimento se ejecuta en el futuro (por ejemplo el próximo lunes). Esta situación ocurriría si, por alguna razón, la respuesta del próximo lunes tuviera un significado especial para nosotros. Los problemas de predicción ocurren con frecuencia en los negocios, donde podríamos estar interesados en las utilidades del siguiente mes de una inversión específica en lugar del promedio de ganancia por inversión en una cartera de valores grande de acciones similares.

Observe que Y es una variable aleatoria, no un parámetro, por lo que predecir su valor representa apartarnos de nuestra meta previa de hacer inferencias acerca de parámetros poblacionales. Si es razonable suponer que ε está distribuido normalmente con media 0 y varianza σ^2, se deduce que Y está distribuida normalmente con media $\beta_0 + \beta_1 x$ y varianza σ^2. Si se conoce la distribución de una variable aleatoria Y y entonces se selecciona un solo valor de Y, ¿cómo se pronosticaría el valor observado? Afirmamos que usted seleccionaría un valor de Y cercano al *centro* de la distribución, en particular un valor cercano al valor esperado de Y. Si estamos interesados en el valor de Y cuando $x = x^*$, emplearíamos $\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*$ como pronosticador de un valor particular de Y^* y también como estimador de $E(Y)$.

Si $x = x^*$, el error de predecir un valor particular de Y^*, usando $\widehat{Y^*}$ como el pronosticador, es la diferencia entre el valor real de Y^* y el valor pronosticado:

$$\text{error} = Y^* - \widehat{Y^*}.$$

Investiguemos ahora las propiedades de este error en el muestreo repetitivo.

Como Y^* y $\widehat{Y^*}$ son variables aleatorias distribuidas normalmente, su diferencia (el error) también está distribuida normalmente.

Al aplicar el Teorema 5.12, que proporciona las fórmulas para el valor esperado y la varianza de una función lineal de variables aleatorias, obtenemos

$$E(\text{error}) = E(Y^* - \widehat{Y^*}) = E(Y^*) - E(\widehat{Y^*}),$$

y como $E(\widehat{Y^*}) = \beta_0 + \beta_1 x^* = E(Y^*)$,

$$E(\text{error}) = 0.$$

Del mismo modo,

$$V(\text{error}) = V(Y^* - \widehat{Y^*}) = V(Y^*) + V(\widehat{Y^*}) - 2\text{Cov}(Y^*, \widehat{Y^*}).$$

Debido a que estamos pronosticando un valor futuro de Y^* que no se emplea en el cálculo de $\widehat{Y^*}$, se deduce que Y^* y $\widehat{Y^*}$ son independientes y por tanto que $\text{Cov}(Y^*, \widehat{Y^*}) = 0$. Entonces,

$$\begin{aligned} V(\text{error}) &= V(Y^*) + V(\widehat{Y^*}) = \sigma^2 + V(\hat{\beta}_0 + \hat{\beta}_1 x^*) \\ &= \sigma^2 + \left(\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}\right)\sigma^2 \\ &= \sigma^2\left[1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}\right]. \end{aligned}$$

Hemos demostrado que el error de pronosticar un valor particular de Y está distribuido normalmente con media 0 y varianza como se da en la ecuación precedente. Se deduce que

$$Z = \frac{Y^* - \widehat{Y^*}}{\sigma\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

tiene una distribución normal estándar. Además, si S es sustituida por σ, se puede demostrar que

$$T = \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

tiene una distribución t de Student con $n - 2$ grados de libertad. Usamos este resultado para calcular el límite en el error de predicción; al hacerlo, construimos un *intervalo de predicción* para la variable aleatoria Y^*. El procedimiento empleado es similar al usado para construir los intervalos de confianza presentados en los capítulos anteriores.

Empezamos por observar que

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha.$$

Sustituyendo por T, obtenemos

$$P\left[-t_{\alpha/2} < \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}} < t_{\alpha/2}\right] = 1 - \alpha.$$

En otras palabras, en un muestreo repetitivo la desigualdad dentro del paréntesis rectangular se cumplirá con una probabilidad igual a $(1 - \alpha)$. Además, la desigualdad continuará cumpliéndose con la misma probabilidad si cada término se multiplica por el mismo factor positivo o si la misma cantidad se suma a cada término de la desigualdad. Multiplique cada término por

$$S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

y luego sume $\widehat{Y^*}$ a cada uno para obtener

$$P\left[\widehat{Y^*} - t_{\alpha/2}S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}} < Y^* < \widehat{Y^*} + t_{\alpha/2}S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}\right] = 1 - \alpha.$$

Entonces, hemos colocado un intervalo alrededor de $\widehat{Y^*}$ que en muestreo repetido contendrá el valor real de Y^* con probabilidad $1 - \alpha$. Esto es, hemos obtenido un intervalo de predicción de $100(1 - \alpha)\%$ para Y^*.

Intervalo de predicción de $100(1 - \alpha)\%$ para Y cuando $x = x^*$

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2}S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}.$$

Al tratar de poner un límite en el error de pronosticar Y, esperaríamos que el error fuera menor en valor absoluto que

$$t_{\alpha/2}S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

con probabilidad igual a $(1 - \alpha)$.

Observe que la longitud de un *intervalo de confianza* para $E(Y)$ cuando $x = x^*$ está dada por

$$2 \times t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}},$$

mientras que la longitud de un *intervalo de predicción para un valor real* de Y cuando $x = x^*$ está dada por

$$2 \times t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}.$$

Entonces, observamos que los *intervalos de predicción* para el valor real de Y son más largos que los *intervalos de confianza* para $E(Y)$ si ambos están determinados por el mismo valor de x^*.

EJEMPLO **11.7** Suponga que el experimento que generó los datos del Ejemplo 11.1 se va a realizar de nuevo con $x = 2$. Pronostique el valor particular de Y con $1 - \alpha = .90$.

Solución Del Ejemplo 11.1, tenemos

$$\hat{\beta}_0 = 1 \quad \text{y} \quad \hat{\beta}_1 = .7,$$

de modo que el valor pronosticado de Y con $x = 2$ es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* = 1 + (.7)(2) = 2.4.$$

Además, con $x^* = 2$,

$$\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}} = \frac{1}{5} + \frac{(2-0)^2}{10} = .6.$$

Del Ejemplo 11.3 sabemos que $s = .606$. El valor $t_{.05}$ con 3 grados de libertad es 2.353. Por tanto, el intervalo de predicción es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

$$2.4 \pm (2.353)(.606)\sqrt{1 + .6}$$

$$2.4 \pm 1.804.$$

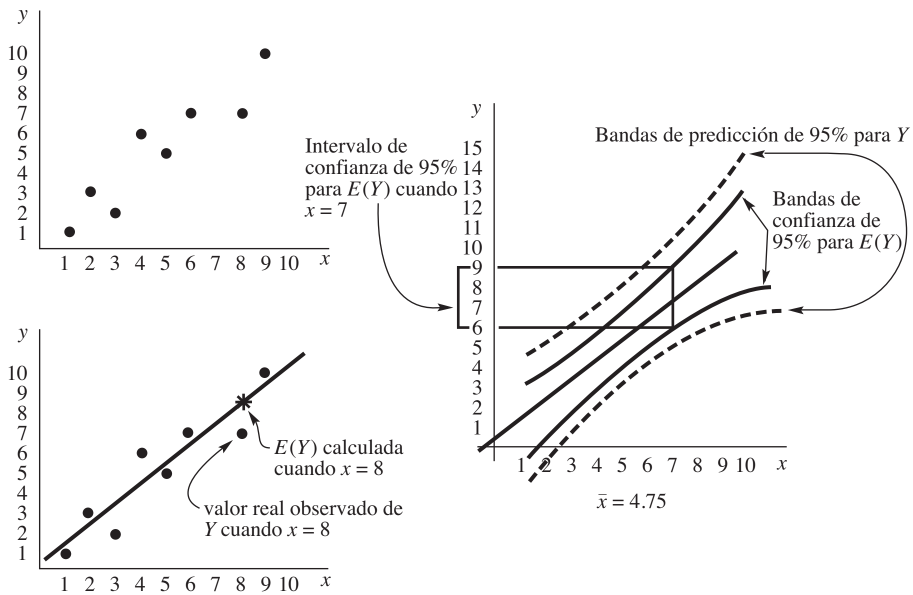
■

La Figura 11.7 representa algunos datos hipotéticos y la recta de regresión estimada, ajustada a la información que indica el valor estimado de $E(Y)$ cuando $x = 8$. También se muestran en la gráfica las *bandas de confianza* para $E(Y)$. Para cada valor de x, calculamos

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}}.$$

Por tanto, para cada valor de x obtenemos un intervalo de confianza para $E(Y)$. El intervalo de confianza para $E(Y)$ cuando $x = 7$ se muestra en el eje y de la figura. Observe que la distancia entre las bandas de confianza es más pequeña cuando $x = \overline{x}$, como se esperaba. Usando el mismo método, calculamos bandas de predicción de un valor Y real para cada ajuste de x. Como ya dijimos antes, para cada valor fijo de x, el intervalo de predicción es más ancho que

FIGURA **11.7** Alguna información hipotética y bandas de confianza y predicción con ella relacionadas



el intervalo de confianza correspondiente. El resultado es que las bandas de predicción caen uniformemente más lejos de la recta de predicción que las bandas de confianza. Las bandas de predicción también están más cercanas entre sí cuando $x = \bar{x}$.

Ejercicios

11.42 Suponga que el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ se ajusta a los n puntos $(y_1, x_1), \ldots, (y_n, x_n)$. ¿A qué valor de x se minimiza la longitud del intervalo de predicción para Y?

11.43 Consulte los Ejercicios 11.5 y 11.17. Use los datos y el modelo dados ahí para construir un intervalo de predicción de 95% para la mediana de los precios de venta en 1980.

11.44 Consulte el Ejercicio 11.43. Encuentre un intervalo de predicción de 95% para la mediana de los precios de venta para el año 1981. Repita para 1982. ¿Consideraría adecuado utilizar este modelo y los datos del Ejercicio 11.5 para predecir la mediana de los precios de venta para el año 1988?

11.45 Consulte los Ejercicios 11.8 y 11.18. Encuentre un intervalo de predicción de 95% para un LC50 de flujo si se observa que el LC50 estático es de 12 partes por millón. Compare la longitud de este intervalo con la del intervalo hallado en el Ejercicio 11.37.

11.46 Consulte el Ejercicio 11.16. Encuentre un intervalo de predicción de 95% para la potencia de una onza de antibiótico almacenado a 65°F. Compare este intervalo con el calculado en el Ejercicio 11.39.

11.47 Consulte el Ejercicio 11.14. Encuentre un intervalo de predicción de 95% para la proporción de sobrevivientes cuando $x = .60$.

11.8 Correlación

Las secciones previas de este capítulo se ocuparon de modelar una respuesta Y como función lineal de una variable no aleatoria x para que pudieran hacerse inferencias apropiadas respecto al valor esperado de Y o un valor futuro de Y para un valor dado de x. Estos modelos son útiles en dos situaciones prácticas bastante diferentes.

Primero, la variable x puede ser controlada completamente por el experimentador. Esto ocurre, por ejemplo, si x es el ajuste de temperatura y Y es el rendimiento en un experimento químico. Entonces, x es sólo el punto en el cual se fija la temperatura marcada cuando el experimento se ejecuta. Desde luego, x podría variar de un experimento a otro, pero bajo el completo control del experimentador, hablando en términos prácticos. El modelo lineal

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

implica entonces que

$$E(Y) = \beta_0 + \beta_1 x$$

o que el rendimiento promedio sea una función lineal del ajuste de temperatura.

En segundo término, la variable x puede ser un valor observado de una variable aleatoria X. Por ejemplo, podríamos relacionar el volumen de madera utilizable Y en un árbol con la circunferencia X de la base. Si pudiera establecerse una relación funcional, entonces en el futuro podríamos predecir la cantidad de madera en cualquier árbol con sólo medir la circunferencia de la base. Para esta situación usamos el modelo

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

para implicar que

$$E(Y|X = x) = \beta_0 + \beta_1 x.$$

Esto es, estamos suponiendo que la esperanza condicional de Y para un valor fijo de X es una función lineal del valor x. Por lo general suponemos que la variable aleatoria vectorial (X, Y) tiene una distribución normal bivariable con $E(X) = \mu_X$, $E(Y) = \mu_Y$, $V(X) = \sigma^2_X$, $V(Y) = \sigma^2_Y$, y el coeficiente de correlación ρ (véase la Sección 5.10), en cuyo caso se puede demostrar que

$$E(Y|X = x) = \beta_0 + \beta_1 x, \qquad \text{donde } \beta_1 = \frac{\sigma_Y}{\sigma_X}\rho.$$

La teoría estadística para hacer inferencias acerca de los parámetros β_0 y β_1 es exactamente igual para estos dos casos, pero las diferencias en interpretación del modelo deben recordarse.

Para el caso donde (X, Y) tiene una distribución bivariable, el experimentador no siempre puede estar interesado en la relación lineal que define $E(Y|X)$. Él o ella pueden querer saber sólo si las variables aleatorias X y Y son *independientes*. Si (X, Y) tiene una distribución normal bivariable (véase la Sección 5.10), entonces la prueba de independencia es equivalente a probar si el coeficiente de correlación ρ es igual a cero. Recuerde de la Sección 5.7 que ρ es positiva si X y Y tienden a aumentar juntas y ρ es negativa si Y disminuye cuando X aumenta.

Denotemos con $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria de una distribución normal bivariante. El estimador de máxima probabilidad de ρ está dado por el coeficiente de correlación muestral:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}.$$

Observe que podemos expresar r en términos de cantidades conocidas:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \hat{\beta}_1\sqrt{\frac{S_{xx}}{S_{yy}}}.$$

Se deduce que r y $\hat{\beta}_1$ tienen el mismo signo.

En el caso donde (X, Y) tenga una distribución normal bivariable, hemos indicado que

$$E(Y|X = x) = \beta_0 + \beta_1 x, \qquad \text{donde } \beta_1 = \frac{\sigma_Y}{\sigma_X}\rho.$$

Entonces, por ejemplo, probar H_0: $\rho = 0$ contra H_a: $\rho > 0$ es equivalente a probar H_0: $\beta_1 = 0$ contra H_a: $\beta_1 > 0$. Del mismo modo, H_a: $\rho < 0$ es equivalente a H_a: $\beta_1 < 0$ y H_a: $\rho \neq 0$ es equivalente a H_a: $\beta_1 \neq 0$. Las pruebas para cada uno de estos conjuntos de hipótesis que contienen β_1 pueden estar basadas (véase la Sección 11.5) en el estadístico

$$t = \frac{\hat{\beta}_1 - 0}{S/\sqrt{S_{xx}}},$$

que tiene una distribución t con $n - 2$ grados de libertad. De hecho (véase el Ejercicio 11.55), este estadístico se puede reescribir en términos de r como sigue:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Debido a que los dos estadísticos t anteriores son equivalentes algebraicos, ambos tienen la misma distribución: la distribución t con $n - 2$ grados de libertad.

Parecería lógico usar r como estadístico de prueba para probar hipótesis más generales acerca de ρ, pero la distribución de probabilidad para r es difícil de obtener. La dificultad puede ser superada, en muestras moderadamente grandes, usando el hecho de que $(1/2)\ln[(1 + r)/(1 - r)]$ está distribuida normalmente en forma aproximada con media $(1/2)\ln[(1 + \rho)/(1 - \rho)]$ y varianza $1/(n - 3)$. Entonces, para probar la hipótesis H_0: $\rho = \rho_0$, podemos emplear una prueba Z en la que

$$Z = \frac{\left(\frac{1}{2}\right)\ln\left(\frac{1+r}{1-r}\right) - \left(\frac{1}{2}\right)\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\frac{1}{\sqrt{n-3}}}.$$

Si α es la probabilidad deseada de cometer un error tipo I, la forma de la región de rechazo depende de la hipótesis alternativa. Las diversas alternativas de interés más frecuente y correspondientes regiones de rechazo son las siguientes:

$$H_a\colon \rho > \rho_0, \qquad \text{RR}: z > z_\alpha,$$
$$H_a\colon \rho < \rho_0, \qquad \text{RR}: z < -z_\alpha,$$
$$H_a\colon \rho \neq \rho_0, \qquad \text{RR}: |z| > z_{\alpha/2}$$

Ilustramos con un ejemplo.

EJEMPLO **11.8** Los datos de la Tabla 11.3 representan una muestra de las calificaciones de un examen de conocimientos matemáticos y las calificaciones de cálculo para diez estudiantes de primer año de universidad seleccionados de manera independiente. Dada esta evidencia, ¿diría usted que las calificaciones del examen de conocimientos matemáticos y las calificaciones de cálculo son independientes? Use $\alpha = .05$. Identifique el correspondiente nivel de significancia alcanzado.

Solución Establezcamos como hipótesis nula que X y Y son independientes; o bien, suponiendo que (X, Y) tenga una distribución normal bivariable, probamos H_0: $\rho = 0$ contra H_a: $\rho \neq 0$. Como estamos concentrados en $\rho = 0$, la prueba puede estar basada en el estadístico $t = (r\sqrt{n-2})/\sqrt{1-r^2}$. Si denotamos con x las calificaciones del examen de conocimientos matemáticos y las calificaciones de cálculo con y, tenemos

$$\sum x_i = 460, \qquad \sum x_i^2 = 23\,634, \qquad S_{xx} = 2474,$$

$$\sum y_i = 760, \qquad \sum y_i^2 = 59\,816, \qquad S_{yy} = 2056,$$

$$\sum x_i y_i = 36\,854, \qquad S_{xy} = 1894.$$

Entonces,

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1894}{\sqrt{(2474)(2056)}} = .8398.$$

El valor del estadístico de prueba es

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(.8398)\sqrt{8}}{\sqrt{1-.7053}} = 4.375.$$

Debido a que t está basada en $n - 2 = 8$ grados de libertad, $t_{\alpha/2} = t_{.025} = 2.306$; el valor observado de nuestro estadístico de prueba se encuentra en la región de rechazo. Por tanto, la evidencia sugiere enfáticamente que las calificaciones del examen de conocimientos matemáticos y las calificaciones de cálculo son dependientes. Observe que $\alpha = .05$ es la probabilidad de que nuestro estadístico de prueba caiga en la región de rechazo cuando H_0 es verdadera. En consecuencia, confiamos razonablemente en que hemos tomado una decisión correcta.

Debido a que estamos poniendo en práctica una prueba de dos colas, el valor $p = 2P(t > 4.375)$. De los valores contenidos en la Tabla 5, Apéndice 3, se deduce que $P(t > 4.375) <$

Tabla 11.3 Datos para el Ejemplo 11.8

Estudiante	Calificación del examen de conocimientos matemáticos	Calificación final de cálculo
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

.005. Entonces el valor $p < 2(.005) = .010$ y para cualquier valor de α mayor que .01 (incluyendo $\alpha = .05$, como se usa en la parte inicial de este análisis), podríamos concluir que $\rho \neq 0$. La aplicación *Student's t Probabilities and Quantiles*, usada con 8 grados de libertad, da ese valor $p = 2P\,(t > 4.375) = 2(.00118) = .00236$, un valor considerablemente menor que el límite superior para el valor p que se obtuvo usando la Tabla 5. ■

Observe que el cuadrado del coeficiente de correlación se presenta en el denominador del estadístico t usado para realizar la prueba de las hipótesis en el Ejercicio 11.8. El estadístico r^2 se denomina *coeficiente de determinación* y tiene una interpretación interesante y útil. Originalmente (Sección 11.3), definimos la suma de cuadrados del error (SSE) como la suma de los cuadrados de las diferencias entre los valores observado y pronosticado de las y_i,

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

Si el modelo de regresión lineal simple se ajusta bien a los datos, las diferencias entre los valores observado y pronosticado son pequeñas, lo cual lleva a un valor pequeño para la SSE. De igual manera, si el modelo de regresión se ajusta mal, la SSE será grande. En el Ejercicio 11.15 usted demostró que una ecuación cómoda desde el punto de vista computacional para SSE es

$$\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy}, \qquad \text{donde } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

Con el uso de esta expresión fue fácil demostrar (Ejercicio 11.15(b)) que $\text{SSE} \leq S_{yy}$. La cantidad $S_{yy} = \sum(y_i - \overline{y})^2$ proporciona una medida de la variación total entre los valores y, pasando por alto las x. De manera alternativa, la SSE mide la variación en los valores y que permanecen sin explicación después de usar las x para ajustar el modelo de regresión lineal simple. Entonces, la razón SSE/S_{yy} da la proporción de la variación total en las y_i que no es explicada por el modelo de regresión lineal.

Observe que el coeficiente de determinación se puede escribir como

$$r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = \left(\frac{S_{xy}}{S_{xx}}\right)\left(\frac{S_{xy}}{S_{yy}}\right) = \left(\frac{\hat{\beta}_1 S_{xy}}{S_{yy}}\right) = \frac{S_{yy} - \text{SSE}}{S_{yy}} = 1 - \frac{\text{SSE}}{S_{yy}}.$$

Así, r^2 se puede interpretar como la *proporción de la variación total en las y_i que es explicada por la variable x en un modelo de regresión lineal simple.*

EJEMPLO **11.9** Consulte el Ejemplo 11.8 donde calculamos el coeficiente de correlación entre las calificaciones del examen de conocimientos matemáticos y las calificaciones finales de cálculo, para diez estudiantes de primer año de universidad seleccionados de manera independiente. Interprete los valores del coeficiente de correlación y el coeficiente de determinación.

Solución En el Ejemplo 11.8 obtuvimos $r = .8398$. Como r es positiva, concluimos que los estudiantes de primer año de universidad con más altas calificaciones en el examen de conocimientos tienden a obtener más altas calificaciones en cálculo. El coeficiente de determinación es $r^2 = (.8398)^2 = .7053$. Entonces, 70.53% de la variación en las calificaciones finales de cálculo se explica al ajustar el modelo lineal simple usando las calificaciones de conocimientos matemáticos como la variable independiente. El modelo de regresión funciona muy bien. ■

Ejercicios

11.48 La siguiente tabla muestra la carga pico de potencia para una planta generadora de energía eléctrica y la temperatura alta diaria para una muestra aleatoria de 10 días. Pruebe la hipótesis de que el coeficiente de correlación poblacional ρ, entre la carga pico de potencia y la temperatura alta, es cero contra la hipótesis alternativa de que sea positivo. Use $\alpha = .05$. Limite o determine el nivel de significancia alcanzado.

Día	Temperatura alta (ºF)	Carga pico
1	95	214
2	82	152
3	90	156
4	81	129
5	99	254
6	100	266
7	93	210
8	95	204
9	93	213
10	87	150

11.49 **Ejercicio Applet** Consulte el Ejemplo 11.1 y el Ejercicio 11.2. Entre a la aplicación *Fitting a Line Using Least Squares.* Los datos que aparecen en la primera gráfica son del Ejemplo 11.1.

a Arrastre la recta azul para obtener una ecuación que visualmente se ajuste bien a los datos. ¿Qué observa acerca de los valores de la SSE y r^2 cuando mejora el ajuste de la recta? ¿Por qué r^2 aumenta cuando la SSE disminuye?

b Haga clic en el botón “Find Best Model” para obtener la recta de mínimos cuadrados. ¿Cuál es el valor de r^2? ¿Cuál es el valor del coeficiente de correlación?

11.50 **Ejercicio Applet** Consulte los Ejercicios 11.5 y 11.6. Los datos del Ejercicio 11.5 aparecen en la gráfica bajo el encabezado “Another Example” en la aplicación *Fitting a Line Using Least Squares.*

a Arrastre la recta azul para obtener una ecuación que visualmente se ajuste bien a los datos. ¿Qué observa acerca del valor de r^2 cuando mejora el ajuste de la recta?

b Haga clic en el botón “Find Best Model” para obtener la recta de mínimos cuadrados. ¿Cuál es el valor de r^2? ¿Cuál es el valor del coeficiente de correlación?

c ¿Por qué el valor de r^2 es mucho más grande que el valor de r^2 obtenido en el Ejercicio 11.49(b) que utilizó los datos del Ejemplo 11.1?

11.51 En el Ejercicio 11.8 los valores LC50 de flujo y estático podrían considerarse variables aleatorias. Usando los datos del Ejercicio 11.8 haga una prueba para ver si la correlación entre valor estático y de flujo difieren significativamente con respecto a cero. Use $\alpha = .01$. Limite o determine el valor p asociado.

11.52 ¿La densidad de plantas de una especie está relacionada con la altitud a la que se recolectan los datos? Denote con Y la densidad de especie y con X la altitud. Un ajuste de un modelo de regresión lineal simple que usó 14 observaciones dio como resultado $\hat{y} = 21.6 - 7.79x$ y $r^2 = .61$.

a ¿Cuál es el valor del coeficiente de correlación r?

b ¿Qué proporción de la variación en las densidades es explicada por el modelo lineal que usa la altitud como la variable independiente?

c ¿Hay suficiente evidencia en la $\alpha = .05$ para indicar que las densidades de plantas disminuyen con un aumento en la altitud?

11.53 El coeficiente de correlación para las estaturas y pesos de diez jugadores ofensivos de futbol americano se determinaron como $r = .8261$.

a ¿Qué porcentaje de la variación en los pesos se explicó por las estaturas de los jugadores?

b ¿Qué porcentaje de la variación en las estaturas se explicó por los pesos de los jugadores?

c ¿Hay suficiente evidencia en el nivel $\alpha = .01$ para decir que estaturas y pesos están correlacionados positivamente?

d **Ejercicio Applet** ¿Cuál es el nivel de significancia alcanzado y que está asociado con la prueba efectuada en el inciso c?

11.54 Suponga que buscamos un estimador intuitivo para

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

a El estimador del método de momentos de $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ es

$$\widehat{\text{Cov}(X, Y)} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Demuestre que los estimadores del método de momentos para las desviaciones estándar de X y Y son

$$\hat{\sigma}_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \qquad \text{y} \qquad \hat{\sigma}_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}.$$

b Sustituya los estimadores por sus parámetros respectivos en la definición de ρ y obtenga el estimador del método de momentos para ρ. Compare su estimador con r, el estimador de máxima probabilidad para ρ presentado en esta sección.

11.55 Considere el modelo de regresión lineal simple basado en la teoría normal. Si estamos interesados en probar $H_0: \beta_1 = 0$ contra varias alternativas, el estadístico

$$T = \frac{\hat{\beta}_1 - 0}{S/\sqrt{S_{xx}}}$$

tiene una distribución t con $n - 2$ grados de libertad si la hipótesis nula es verdadera. Demuestre que la ecuación para T también se puede escribir como

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

11.56 Consulte el Ejercicio 11.55. ¿Es $r = .8$ suficientemente grande para decir que $\rho > 0$ en el nivel de significancia $\alpha = .05$?

a Suponga $n = 5$ y realice la prueba.

b Suponga $n = 12$ y realice la prueba.

c **Ejercicio Applet** Determine los valores p para las pruebas que llevó a cabo en los incisos a y b.

d ¿Llegó usted a las mismas conclusiones en los incisos a y b? ¿Por qué sí o por qué no?

e ¿Por qué el valor p asociado con la prueba del inciso b es mucho menor que el valor p asociado con la prueba efectuada en el inciso a?

11.57 Consulte los Ejercicios 11.55 y 11.56.

a ¿Qué término en el estadístico T determina si el valor de t es positivo o negativo?

b ¿Qué cantidades determinan el tamaño de $|t|$?

11.58 Consulte el Ejercicio 11.55. Si $n = 4$, ¿cuál es el mínimo valor de r que permitirá concluir que $\rho > 0$ en el nivel de significancia $\alpha = .05$?

11.59 Consulte los Ejercicios 11.55 y 11.58. Si $n = 20$, ¿cuál es el máximo valor de r que permitirá concluir que $\rho < 0$ en el nivel de significancia $\alpha = .05$?

***11.60** Consulte los Ejercicios 11.8 y 11.51. Suponga que pruebas independientes, con los mismos tóxicos y especies pero en diferente laboratorio, demostraron que $r = .85$ con $n = 20$. Pruebe la hipótesis de que los dos coeficientes de correlación entre mediciones estática y de flujo de LC50 son iguales. Use $\alpha = .05$.

11.9 Algunos ejemplos prácticos

En esta sección presentamos dos ejemplos que ilustran la aplicabilidad de técnicas previamente desarrolladas a datos reales. Casi todos los métodos están ilustrados en algún punto en el curso del análisis. No intentamos aplicar todos los métodos para cada ejemplo.

EJEMPLO **11.10** En su tesis de doctorado, H. Behbahani examinó el efecto de hacer variar la proporción de agua y cemento en la resistencia del concreto después de 28 días. Para el concreto con un contenido de cemento de 200 libras por yarda cúbica, obtuvo los datos presentados en la Tabla 11.4.[9] Sea Y la resistencia y x la proporción de agua y cemento.

a Ajuste el modelo $E(Y) = \beta_0 + \beta_1 x$.
b Pruebe $H_0 : \beta_1 = 0$ contra $H_a : \beta_1 < 0$ con $\alpha = .05$. (Observe que si H_0 es rechazada concluimos que $\beta_1 < 0$ y que la resistencia tiende a disminuir con un aumento en la proporción de agua y cemento.) Identifique el correspondiente nivel de significancia alcanzado.
c Encuentre un intervalo de confianza de 90% para la resistencia esperada del concreto cuando la proporción de agua y cemento sea de 1.5. ¿Qué le ocurrirá al intervalo de confianza si tratamos de calcular resistencias medias para proporciones de agua y cemento de .3 o 2.7?

Solución **a** Usando las fórmulas desarrolladas en la Sección 11.3 obtenemos

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = 8.709 - \frac{1}{6}(8.74)(6.148) = -.247,$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 12.965 - \frac{1}{6}(8.74)^2 = .234,$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 6.569 - \frac{1}{6}(6.148)^2 = .269,$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-0.247}{0.234} = -1.056,$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \frac{6.148}{6} - (-1.056)\left(\frac{8.74}{6}\right) = 2.563.$$

(En todo este ejemplo, todos los cálculos se hacen a tres lugares decimales.)

9. *Fuente*: datos adaptados de la obra de Hamid Behbahani, "Econocrete—Design and Properties" (tesis para Ph.D., University of Florida, 1977), p. 95.

Tabla 11.4 Datos para el Ejemplo 11.10

Proporción de agua y cemento	Resistencia (100 ft/lb)
1.21	1.302
1.29	1.231
1.37	1.061
1.46	1.040
1.62	.803
1.79	.711

Por tanto, el modelo de línea recta que mejor se ajusta a los datos es

$$\hat{y} = 2.563 - 1.056x.$$

b Debido a que deseamos probar si hay evidencia de que $\beta_1 < 0$ con $\alpha = .05$, el estadístico de prueba apropiado es

$$t = \frac{\hat{\beta}_1 - 0}{S\sqrt{c_{11}}}, \quad \text{o} \quad t = \frac{\hat{\beta}_1 - 0}{S\sqrt{\dfrac{1}{S_{xx}}}}.$$

Para este modelo de regresión lineal simple,

$$\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy} = .269 - (-1.056)(-.247) = .008,$$

y, por tanto,

$$s = \sqrt{s^2} = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{.008}{4}} = .045.$$

Entonces, el valor del estadístico de prueba apropiado para probar $H_0: \beta_1 = 0$ contra H_a: $\beta_1 < 0$ es

$$t = \frac{-1.056 - 0}{.045\sqrt{1/(.234)}} = -11.355.$$

Como este estadístico está basado en $n - 2$ grados de libertad y la región de rechazo apropiada es $t < -t_{.05} = -2.132$, rechazamos H_0 a favor de H_a en un nivel de significancia de $\alpha = .05$. La prueba apropiada es una prueba de cola inferior, y valor $p = P\,(t < -11.355)$, donde t tiene una distribución t con 4 grados de libertad. La Tabla 5, Apéndice 3, se aplica para dar un valor $p < .005$. De hecho la aplicación *Student't Probabilities and Quantiles* proporciona un valor $p = P\,(t < -11.355) = P\,(t > 11.355) = .00017$, valor considerablemente menor que .005. En consecuencia, para los valores de α que más se usan, concluimos que hay evidencia para indicar que la resistencia disminuye con un aumento en la proporción de agua y cemento *en la región donde el experimento fue realizado*. Desde un punto de vista práctico, la proporción de agua y cemento debe ser lo suficientemente grande para humedecer el cemento, la arena y otros componentes que conforman el concreto. Pero si la proporción de agua y cemento se hace demasiado grande, el concreto será inútil.

c Debido a que estamos usando un modelo de regresión lineal simple, el intervalo de confianza puede obtenerse de la fórmula

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}.$$

Buscamos un intervalo de confianza cuando $x = 1.5$; por tanto, $x^* = 1.5$ y

$$\hat{\beta}_0 + \hat{\beta}_1 x^* = 2.563 - (1.056)(1.5) = .979.$$

Usando cálculos de los incisos a y b, obtenemos el intervalo de confianza de 90% deseado:

$$.979 \pm (2.132)(.045)\sqrt{\frac{1}{6} + \frac{(1.5 - 1.457)^2}{.234}} \quad \text{o} \quad (.938, 1.020).$$

Entonces, estimaríamos que la resistencia media del concreto con una proporción de agua y cemento de 1.5 estaría entre .938 y 1.020.

De la expresión de varianza podemos ver que el intervalo de confianza se hace más ancho cuando x^* se aleja de $\overline{x} = 1.457$. También, los valores $x^* = .3$ y $x^* = 2.7$ están lejos de los valores que se usaron en el experimento. Debe tenerse sumo cuidado antes de construir un intervalo de confianza para $E(Y)$ cuando los valores de x^* están muy lejos de la región experimental. Es muy probable que las proporciones de agua y cemento de .3 y 2.7 produzcan un concreto que sea inútil por completo. ■

En muchas situaciones prácticas, el componente determinístico más apropiado de un modelo no es lineal. Por ejemplo, un gran número de poblaciones de plantas o animales tienden a crecer a ritmos exponenciales. Si Y_t denota el tamaño de la población en el tiempo t, podríamos emplear el modelo

$$E(Y_t) = \alpha_0 e^{\alpha_1 t}.$$

Aun cuando esta expresión no es lineal en los parámetros α_0 y α_1, se puede hacer lineal si se aplican logaritmos naturales. Si Y_t se puede observar para diversos valores de t, podemos escribir el modelo como

$$\ln Y_t = \ln \alpha_0 + \alpha_1 t + \varepsilon$$

y estimar $\ln \alpha_0$ y α_1 por el método de mínimos cuadrados.

Otros modelos básicos también se pueden hacer lineales. En las ciencias biológicas en ocasiones es posible relacionar el peso (o volumen) de un organismo con respecto a alguna medición lineal como la longitud (o el peso). Si W denota el peso y l la longitud, a menudo se aplica el modelo

$$E(W) = \alpha_0 l^{\alpha_1}$$

para α_0 y α_1 desconocidas. (Este modelo se conoce como *ecuación alométrica*.) Si deseamos relacionar el peso de organismos seleccionados al azar con longitudes fijas observables, podemos aplicar logaritmos y obtener el modelo lineal

$$\ln W = \ln \alpha_0 + \alpha_1 \ln l + \varepsilon = \beta_0 + \beta_1 x + \varepsilon$$

con $x = \ln l$. Entonces, $\beta_0 = \ln \alpha_0$ y $\beta_1 = \alpha_1$ se pueden estimar por el método de mínimos cuadrados. El siguiente ejemplo ilustra ese modelo.

EJEMPLO **11.11** En el conjunto de datos de la Tabla 11.5, W denota el peso (en libras) y l la longitud (en pulgadas) de 15 lagartos capturados en la región central de Florida. Debido a que l es más fácil de observar (quizá de una fotografía) que W para lagartos en su hábitat natural, buscamos

Tabla 11.5 Datos para el Ejemplo 11.11

Lagarto	$x = \ln l$	$y = \ln W$
1	3.87	4.87
2	3.61	3.93
3	4.33	6.46
4	3.43	3.33
5	3.81	4.38
6	3.83	4.70
7	3.46	3.50
8	3.76	4.50
9	3.50	3.58
10	3.58	3.64
11	4.19	5.90
12	3.78	4.43
13	3.71	4.38
14	3.73	4.42
15	3.78	4.25

construir un modelo que relacione peso con longitud. Ese modelo se puede usar entonces para predecir los pesos de lagartos de longitudes especificadas. Ajuste el modelo

$$\ln W = \ln \alpha_0 + \alpha_1 \ln l + \varepsilon = \beta_0 + \beta_1 x + \varepsilon$$

a los datos. Encuentre un intervalo de predicción de 90% para W si se observa que $\ln l$ es 4.00.

Solución Empezamos por calcular las cantidades que tienen aplicación de rutina en toda nuestra solución:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = 251.9757 - \frac{1}{15}(56.37)(66.27) = 2.933,$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2 = 212.6933 - \frac{1}{15}(56.37)^2 = 0.8548,$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} y_i \right)^2 = 303.0409 - \frac{1}{15}(66.27)^2 = 10.26,$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{2.933}{0.8548} = 3.4312,$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \frac{66.27}{15} - (3.4312)\left(\frac{56.37}{15}\right) = -8.476.$$

Podemos ahora calcular α_0 con

$$\hat{\alpha}_0 = e^{\hat{\beta}_0} = e^{-8.476} = .0002$$

y α_1 con $\hat{\alpha}_1 = \hat{\beta}_1$ para llegar al modelo estimado

$$\hat{w} = \hat{\alpha}_0 l^{\hat{\alpha}_1} = (.0002) l^{3.4312}.$$

(En muchos casos α_1 será cercano a 3 porque peso o volumen es aproximadamente proporcional al cubo de una medición lineal.)

Para estos datos, SSE $= .1963$, $n = 15$ y $s = \sqrt{\text{SSE}/(n-2)} = .123$. Los cálculos que llevan a esos valores numéricos son análogos por completo a los del Ejemplo 11.10.

Para hallar un intervalo de predicción para W, donde $x = \ln l = 4$, primero debemos formar un intervalo de predicción para $Y = \ln W$. Como antes, el intervalo de predicción es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{.05} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}},$$

donde $t_{.05}$ está basada en $n - 2 = 13$ grados de libertad. Por tanto, $t_{.05} = 1.771$ y el intervalo de predicción de 90% para $Y = \ln W$ es

$$-8.476 + 3.4312(4) \pm 1.771(.123)\sqrt{1 + \frac{1}{15} + \frac{(4 - 3.758)^2}{.8548}}$$

$$5.2488 \pm .2321,$$

o bien,

$$(5.0167, 5.4809).$$

Como $\hat{Y} = \ln \hat{W}$, podemos predecir W con $e^{\hat{Y}} = e^{5.2488} = 190.3377$. El intervalo de predicción de 90% observado para W es

$$\left(e^{5.0167}, e^{5.4809}\right), \quad \text{o} \quad (150.9125,\ 240.0627).$$

Cuando $x = \ln l = 4$, entonces $l = e^4 = 54.598$. Entonces, para un lagarto de 54.598 pulgadas de largo, predecimos que su peso estará entre 150.91 y 240.06 libras. El intervalo relativamente estrecho en la escala de logaritmos naturales se hace un intervalo más bien ancho cuando se transforma a la escala original. ■

Los datos presentados y analizados en esta sección son ejemplos tomados de experimentos reales; los métodos desarrollados en secciones previas de este capítulo se aplicaron para producir respuestas de interés real para experimentadores. A lo largo del Ejemplo 11.11 hemos demostrado la forma en que la teoría de modelos lineales a veces se puede aplicar después de la transformación de la escala de las variables originales. Desde luego que no todos los modelos se pueden hacer lineales, pero existen numerosas técnicas para calcular mínimos cuadrados no lineales.

Ejercicios

11.61 Consulte el Ejemplo 11.10. Encuentre un intervalo de predicción de 90% para la resistencia del concreto cuando la proporción de agua y cemento sea 1.5.

11.62 Consulte el Ejercicio 11.11. Calcule el coeficiente de correlación r entre las variables $\ln W$ y $\ln l$. ¿Qué proporción de la variación en $y = \ln w$ está explicada por $x = \ln l$?

***11.63** Es bien sabido que grandes cuerpos de agua tienen un efecto mitigador en la temperatura de las masas de tierra circundantes. En una noche fría en la región central de Florida se registraron temperaturas a distancias iguales a lo largo de una zona que corría a favor del viento desde un lago de grandes dimensiones. Los datos resultantes se dan en la siguiente tabla.

Lugar (x)	Temperatura °F, (y)
1	37.00
2	36.25
3	35.41
4	34.92
5	34.52
6	34.45
7	34.40
8	34.00
9	33.62
10	33.90

Observe que las temperaturas bajan rápidamente y luego se nivelan cuando nos alejamos del lago. El modelo sugerido para estos datos es

$$E(Y) = \alpha_0 e^{-\alpha_1 x}.$$

a Convierta el modelo en uno lineal y calcule los parámetros por medio del método de mínimos cuadrados.

b Encuentre un intervalo de confianza de 90% para α_0. Interprete el resultado.

***11.64** Consulte el Ejercicio 11.14. Un modelo propuesto para estos datos sobre la proporción de sobrevivientes a la contaminación térmica es

$$E(Y) = \exp(-\alpha_0 x^{\alpha_1}).$$

Linealice este modelo y estime los parámetros con el método de mínimos cuadrados y los datos del ejercicio 11.14 (ignore la observación con $y = 1.00$)

***11.65** En las ciencias biológicas y fisicas un modelo común para el crecimiento proporcional en el tiempo es:

$$E(Y) = 1 - e^{-\beta t},$$

donde Y denota una proporción y t denota tiempo. Y podría representar la proporción de huevos que empollan, la proporción de un organismo lleno de células enfermas, la proporción de pacientes que reaccionan a un medicamento o la proporción de un líquido que ha pasado por un medio poroso. Con n observaciones de la forma (y_i, t_i), haga un resumen de la forma en que calcularía y luego formaría un intervalo de confianza para β.

11.10 Ajuste del modelo lineal mediante matrices

Hasta este punto en este capítulo hemos trabajado en forma casi exclusiva con modelos de regresión lineal *simple*, que han hecho posible que expresemos nuestras deduccciones y resultados usando expresiones algebraicas ordinarias. La única forma práctica de manejar resultados y deducciones análogos para modelos de regresión lineal *múltiple* es por medio de álgebra de matrices. En esta sección usamos matrices para expresar de otro modo algunos de los resultados anteriores y ampliar estos resultados al modelo de regresión lineal múltiple.

Suponga que tenemos el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

y hacemos n observaciones independientes, $y_1, y_2, \ldots, y_n$, en Y. Podemos escribir la observación y_i como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

donde x_{ij} es el ajuste de la j-ésima variable independiente para la i-ésima observación, $i = 1, 2, \ldots, n$. Ahora definimos las siguientes matrices con $x_0 = 1$:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} x_0 & x_{11} & x_{12} & \cdots & x_{1k} \\ x_0 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ x_0 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Entonces, las n ecuaciones que representan y_i como función de las x, las β y las ε se pueden escribir simultáneamente como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

(En el Apéndice 1 véase una exposición de las operaciones con matrices.)

Para n observaciones desde un modelo lineal simple de la forma

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

tenemos

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

(Suprimimos el segundo subíndice en x porque sólo aparece una variable x.) Las ecuaciones de mínimos cuadrados para β_0 y β_1 se dieron en la Sección 11.3 como

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i,$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

Como

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

y

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix},$$

si

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

vemos que las ecuaciones de mínimos cuadrados están dadas por

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{Y}.$$

Por tanto,

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Aun cuando hemos demostrado que este resultado se cumple para un caso simple, se puede demostrar que en general las ecuaciones de mínimos cuadrados y las soluciones presentadas en notación matricial son las siguientes.

Ecuaciones de mínimos cuadrados y soluciones para un modelo lineal general

Ecuaciones: $(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{Y}$.

Soluciones: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

EJEMPLO **11.12** Resuelva el Ejemplo 11.1 usando operaciones matriciales.

Solución De los datos dados en el Ejemplo 11.1 vemos que

$$\mathbf{Y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix}, \qquad \text{y} \qquad \mathbf{X} = \begin{array}{c} \begin{array}{cc} x_0 & x_1 \end{array} \\ \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \end{array}.$$

Se deduce que

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}, \qquad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \qquad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Por tanto,

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ .7 \end{bmatrix},$$

o $\hat{\beta}_0 = 1$ y $\hat{\beta}_1 = .7$. Entonces,

$$\hat{y} = 1 + .7x,$$

igual que en el Ejemplo 11.1. ■

EJEMPLO 11.13 Ajuste una parábola a los datos del Ejemplo 11.1 usando el modelo

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon.$$

Solución La matriz **X** para este ejemplo difiere de la del Ejemplo 11.12 sólo por la adición de una tercera columna correspondiente a x^2. (Observe que $x_1 = x$, $x_2 = x^2$ y $k = 2$ en la notación del modelo lineal general.). Entonces,

$$\mathbf{Y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix}, \qquad \mathbf{X} = \begin{array}{c} \begin{array}{ccc} x_0 & x & x^2 \end{array} \\ \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \end{array}.$$

(Las tres variables, x_0, x y x^2, se muestran arriba de sus columnas respectivas en la matriz **X**.) Por tanto, para la primera medición, $y = 0$, $x_0 = 1$, $x = -2$ y $x^2 = 4$; y para la segunda medición, $y = 0$, $x_0 = 1$, $x = -1$ y $x^2 = 1$. Las filas sucesivas de las matrices **Y** y **X** se obtienen de un modo semejante.

Los productos matriciales **X′X** y **X′Y** son

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix},$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 13 \end{bmatrix}.$$

Omitimos el proceso de invertir **X′X** y simplemente expresamos que la matriz inversa es igual a

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 17/35 & 0 & -1/7 \\ 0 & 1/10 & 0 \\ -1/7 & 0 & 1/14 \end{bmatrix}.$$

[Usted puede verificar que $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$.]

Finalmente,

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \begin{bmatrix} 17/35 & 0 & -1/7 \\ 0 & 1/10 & 0 \\ -1/7 & 0 & 1/14 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \\ 13 \end{bmatrix} = \begin{bmatrix} 4/7 \\ 7/10 \\ 3/14 \end{bmatrix} \approx \begin{bmatrix} .571 \\ .700 \\ .214 \end{bmatrix}. \end{aligned}$$

Por tanto, $\hat{\beta}_0 = .571$, $\hat{\beta}_1 = .7$ y $\hat{\beta}_2 = .214$, y la ecuación de predicción es

$$\hat{y} = .571 + .7x + .214x^2.$$

Una gráfica de esta parábola en la Figura 11.6 indicará un buen ajuste para los puntos. ■

Las expresiones para $V(\hat{\beta}_0)$, $V(\hat{\beta}_1)$, $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$, así como SSE que dedujimos en la Sección 11.4 para el modelo de regresión lineal *simple*, se pueden expresar de manera conveniente en términos de matrices. Hemos visto que para el modelo lineal $Y = \beta_0 + \beta_1 x + \varepsilon$, $\mathbf{X}'\mathbf{X}$ está dado por

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}.$$

Se puede demostrar que

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \dfrac{\sum x_i^2}{nS_{xx}} & -\dfrac{\overline{x}}{S_{xx}} \\ -\dfrac{\overline{x}}{S_{xx}} & \dfrac{1}{S_{xx}} \end{bmatrix} = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}.$$

Si se verifican las varianzas y las covarianzas derivadas en la Sección 11.4, podemos ver que

$$V(\hat{\beta}_i) = c_{ii}\sigma^2, \qquad i = 0, 1$$

y

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2 = c_{10}\sigma^2.$$

Recuerde que un estimador insesgado para σ^2, la varianza del término de error ε, está dado por $S^2 = \text{wSSE}/(n - 2)$. Un poco de álgebra de matrices demostrará que $\text{SSE} = \sum(y_i - \hat{y}_i)^2$ se puede expresar como

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}.$$

(Observe que $\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$.)

EJEMPLO **11.14** Determine las varianzas de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ para el Ejemplo 11.12 y proponga un estimador para σ^2.

Solución En el Ejemplo 11.12 encontramos que

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Por tanto,

$$V(\hat{\beta}_0) = c_{00}\sigma^2 = (1/5)\sigma^2,$$
$$V(\hat{\beta}_1) = c_{11}\sigma^2 = (1/10)\sigma^2.$$

Al igual que antes, $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 0$ en este caso porque $\sum x_i = 0$. Para estos datos,

$$\mathbf{Y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \hat{\beta} = \begin{bmatrix} 1 \\ .7 \end{bmatrix}.$$

En consecuencia,

$$\begin{aligned}
\text{SSE} &= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} \\
&= [0 \quad 0 \quad 1 \quad 1 \quad 3]\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix} - [1 \quad .7]\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix} \\
&= 11 - [1 \quad .7]\begin{bmatrix} 5 \\ 7 \end{bmatrix} = 11 - 9.9 = 1.1.
\end{aligned}$$

Entonces,

$$s^2 = \frac{\text{SSE}}{n-2} = \frac{1.1}{5-2} = \frac{1.1}{3} = .367.$$

Observe la correspondencia con los resultados que se obtuvieron en los Ejemplos 11.2 y 11.3. ■

Ejercicios

11.66 Consulte el Ejercicio 11.3. Ajuste el modelo sugerido aplicando matrices.

11.67 Use el método de matrices para ajustar una recta a los datos de la siguiente tabla, grafique los puntos y luego trace la recta ajustada como prueba de los cálculos. Los datos son los mismos de los Ejercicios 11.3 y 11.66 excepto que están recorridos 1 unidad en la dirección positiva a lo largo del eje x. ¿Qué efecto tiene la separación simétrica en los valores x alrededor de $x = 0$ sobre la forma de la matriz $(\mathbf{X}'\mathbf{X})$ y los cálculos resultantes?

y	x
3	−1
2	0
1	1
1	2
.5	3

11.68 Ajuste el modelo cuadrático $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ a los datos de la tabla siguiente. Localice los puntos y trace la parábola ajustada como prueba de los cálculos.

y	x
1	−3
0	−2
0	−1
−1	0
−1	1
0	2
0	3

11.69 El fabricante de autos Lexus ha aumentado continuamente sus ventas desde el lanzamiento de esa marca en 1989 en Estados Unidos. No obstante, el porcentaje de aumento cambió en 1996 cuando el Lexus introdujo una línea de camiones. Las ventas de vehículos Lexus de 1996 a 2003 se muestran en la siguiente tabla.[10]

x	y
1996	18.5
1997	22.6
1998	27.2
1999	31.2
2000	33.0
2001	44.9
2002	49.4
2003	35.0

a Denotando con Y las ventas y con x el año cifrado (–7 para 1996, –5 para 1997, hasta 7 para 2003), ajuste el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$.

b Para los mismos datos, ajuste el modelo $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$.

11.70 **a** Calcule SSE y S^2 para el Ejercicio 11.4. Use el método de matrices.

b Ajuste el modelo sugerido en el Ejercicio 11.4 para la relación entre valores auditados y valores en libros usando matrices. Podemos simplificar los cálculos si definimos

$$x_i^* = x_i - \overline{x}$$

y ajustando el modelo $Y = \beta_0^* + \beta_1^* x^* + \varepsilon$. Adecue este último modelo y calcule la SSE. Compare su respuesta con el cálculo de la SSE del inciso a.

11.11 Funciones lineales de los parámetros del modelo: regresión lineal múltiple

Todos los resultados teóricos de la Sección 11.4 se pueden ampliar al modelo de regresión lineal múltiple,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n.$$

Suponga que $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ son variables aleatorias independientes con $E(\varepsilon_i) = 0$ y $V(\varepsilon_i) = \sigma^2$. Entonces los estimadores de mínimos cuadrados están dados por

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

siempre que $(\mathbf{X}'\mathbf{X})^{-1}$ exista. Las propiedades de estos estimadores son las siguientes (se omite la prueba).

10. *Fuente*: adaptado de *Automotive News*, 26 de enero de 2004.

Propiedades de los estimadores de mínimos cuadrados: regresión lineal múltiple

1. $E(\hat{\beta}_i) = \beta_i$, $i = 0, 1, \ldots, k$.
2. $V(\hat{\beta}_i) = c_{ii}\sigma^2$, donde c_{ii} es el elemento en la fila i y columna i de $(\mathbf{X}'\mathbf{X})^{-1}$. (Recuerde que esta matriz tiene una fila y una columna numerados con 0.)
3. $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\sigma^2$, donde c_{ij} es el elemento en la fila i y columna j de $(\mathbf{X}'\mathbf{X})^{-1}$.
4. Un estimador insesgado de σ^2 es $S^2 = \text{SSE}/[n - (k + 1)]$, donde $\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$. (Observe que hay $k + 1$ valores β_i desconocidas en el modelo.)

Si, además, las ε_i, para $i = 1, 2, \ldots, n$ están distribuidas normalmente,

5. Cada $\hat{\beta}_i$ está distribuida normalmente.
6. La variable aleatoria
$$\frac{[n-(k+1)]S^2}{\sigma^2}$$
tiene una distribución χ^2 con $n - (k + 1)$ grados de libertad.
7. Los estadístico S^2 y $\hat{\beta}_i$ son independientes para cada $i = 0, 1, 2, \ldots, k$.

11.12 Inferencias respecto a funciones lineales de los parámetros del modelo: regresión lineal múltiple

Como lo vimos en las Secciones 11.5 y 11.6, podríamos estar interesados en hacer inferencias acerca de una β_i individual o de combinaciones lineales de los parámetros del modelo β_0, $\beta_1, \ldots, \beta_k$. Por ejemplo, queremos estimar $E(Y)$, dada por

$$E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k,$$

donde $E(Y)$ representa la producción media obtenida en un proceso químico para arreglos de variables de proceso controladas $x_1, x_2, \ldots, x_k$; o la utilidad media de una empresa para diversos gastos de inversión $x_1, x_2, \ldots, x_k$. Las propiedades de los estimadores de estas funciones lineales se dan en esta sección.

Suponga que deseamos hacer una inferencia acerca de la función lineal

$$a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k,$$

donde $a_0, a_1, a_2, \ldots, a_k$ son constantes (algunas de las cuales pueden ser iguales a cero). Definiendo la matriz $(k + 1) \times 1$,

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix},$$

se deduce que una combinación lineal de las $\beta_0, \beta_1, \ldots, \beta_k$ correspondiente a $a_0, a_1, \ldots, a_k$ se puede expresar como

$$\mathbf{a}'\beta = a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k.$$

De aquí en adelante nos referiremos a esas combinaciones lineales en su forma matricial. Como $\mathbf{a}'\beta$ es una combinación lineal de los parámetros del modelo, un estimador insesgado para $\mathbf{a}'\beta$ está dado por la misma combinación lineal de los estimadores paramétricos. Esto es, por el Teorema 5.12, si

$$\widehat{\mathbf{a}'\beta} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k = \mathbf{a}'\hat{\beta},$$

entonces

$$\begin{aligned} E(\mathbf{a}'\hat{\beta}) &= E(a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k) \\ &= a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k = \mathbf{a}'\beta. \end{aligned}$$

Aplicando el mismo teorema, encontramos la varianza de $\mathbf{a}'\hat{\beta}$:

$$\begin{aligned} V(\mathbf{a}'\hat{\beta}) &= V(a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k) \\ &= a_0^2V(\hat{\beta}_0) + a_1^2V(\hat{\beta}_1) + a_2^2V(\hat{\beta}_2) + \cdots + a_k^2V(\hat{\beta}_k) \\ &\quad + 2a_0a_1\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) + 2a_0a_2\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ &\quad + \cdots + 2a_1a_2\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) + \cdots + 2a_{k-1}a_k\mathrm{Cov}(\hat{\beta}_{k-1}, \hat{\beta}_k), \end{aligned}$$

donde $V(\hat{\beta}_i) = c_{ii}\sigma^2$ y $\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\sigma^2$. Usted puede verificar que $V(\mathbf{a}'\hat{\beta})$ está dada por

$$V(\mathbf{a}'\hat{\beta}) = [\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]\sigma^2.$$

Por último, recordando que $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ están distribuidas normalmente en muestreo repetitivo (Sección 11.11), es evidente que $\mathbf{a}'\hat{\beta}$ es una función lineal de variables aleatorias distribuidas normalmente y por tanto ella misma está distribuida normalmente en muestreo repetitivo.

Debido a que $\mathbf{a}'\hat{\beta}$ está distribuida normalmente con

$$E(\mathbf{a}'\hat{\beta}) = \mathbf{a}'\beta$$

y $V(\mathbf{a}'\hat{\beta}) = [\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]\sigma^2$, concluimos que

$$Z = \frac{\mathbf{a}'\hat{\beta} - \mathbf{a}'\beta}{\sqrt{V(\mathbf{a}'\hat{\beta})}} = \frac{\mathbf{a}'\hat{\beta} - \mathbf{a}'\beta}{\sigma\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

tiene una distribución normal estándar y podría emplearse para probar una hipótesis

$$H_0: \mathbf{a}'\beta = (\mathbf{a}'\beta)_0$$

cuando $(\mathbf{a}'\beta)_0$ es algún valor especificado. Del mismo modo, un intervalo de confianza de $100(1 - \alpha)\%$ para $\mathbf{a}'\beta$ es

$$\mathbf{a}'\hat{\beta} \pm z_{\alpha/2}\sigma\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

Además, como podríamos sospechar, si sustituimos S por σ, la cantidad

$$T = \frac{\mathbf{a}'\hat{\beta} - \mathbf{a}'\beta}{S\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

tiene una distribución t de Student en muestreo repetitivo, con $[n - (k + 1)]$ grados de libertad, y proporciona un estadístico de prueba para comprobar la hipótesis

$$H_0 : \mathbf{a}'\beta = (\mathbf{a}'\beta)_0.$$

Una prueba para a′β

$H_0 : \mathbf{a}'\beta = (\mathbf{a}'\beta)_0.$

$$H_a : \begin{cases} \mathbf{a}'\beta > (\mathbf{a}'\beta)_0, \\ \mathbf{a}'\beta < (\mathbf{a}'\beta)_0, \\ \mathbf{a}'\beta \neq (\mathbf{a}'\beta)_0. \end{cases}$$

Estadistico de prueba: $T = \dfrac{\mathbf{a}'\hat{\beta} - (\mathbf{a}'\beta)_{\mathbf{0}}}{S\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}.$

$$\text{Región de rechazo: } \begin{cases} t > t_\alpha, \\ t < -t_\alpha, \\ |t| > t_{\alpha/2}. \end{cases}$$

Aquí, t_α está basada en $[n - (k + 1)]$ grados de libertad.

El correspondiente intervalo de confianza $100(1 - \alpha)\%$ para $\mathbf{a}'\beta$ es el siguiente.

Un intervalo de confianza 100(1 – α)% para a′β

$$\mathbf{a}'\hat{\beta} \pm t_{\alpha/2} S\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

Como ya vimos antes, la $t_{\alpha/2}$ tabulada en esta fórmula está basada en $[n - (k + 1)]$ grados de libertad.

Aunque por lo general no consideramos una β_i individual como una combinación lineal de $\beta_0, \beta_1, \ldots, \beta_k$, si escogemos

$$a_j = \begin{cases} 1, & \text{si } j = i, \\ 0, & \text{si } j \neq i, \end{cases}$$

entonces $\beta_i = \mathbf{a}'\beta$ para esta selección de $\mathbf{a}$. En el Ejercicio 11.71 usted demostrará que con esta selección de $\mathbf{a}$, $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = c_{ii}$, donde c_{ii} es el elemento en la fila i y la columna i de $(\mathbf{X}'\mathbf{X})^{-1}$. Este hecho simplifica enormemente la forma del estadístico de prueba y de los intervalos de confianza que se puedan usar para hacer inferencias acerca de una β_i individual.

Como ya antes dijimos, una aplicación útil de las técnicas de prueba de hipótesis e intervalo de confianza que acabamos de presentar, es al problema de calcular el valor medio de Y, $E(Y)$ para valores fijos de las variables independientes $x_1, x_2, \ldots, x_k$. En particular, si x_i^* denota un valor específico de x_i, para $i = 1, 2, \ldots, k$, entonces

$$E(Y) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_k x_k^*.$$

Observe que $E(Y)$ es un caso especial de $a_0\beta_0 + a_1\beta_1 + \cdots + a_k\beta_k = \mathbf{a}'\beta$ con $a_0 = 1$ y $a_i = x_i^*$, para $i = 1, 2, \ldots, k$. Entonces, una inferencia alrededor de $E(Y)$ cuando $x_i = x_i^*$, para $i = 1, 2, \ldots, k$, se puede hacer usando las técnicas desarrolladas para combinaciones lineales generales de las β.

Ilustramos esto con dos ejemplos.

EJEMPLO **11.15** ¿Los datos del Ejemplo 11.1 presentan suficiente evidencia para indicar curvatura en la función de respuesta? Pruebe usando $\alpha = .05$ y obtenga límites para el nivel de significancia alcanzado.

Solución La pregunta anterior supone que el modelo probabilístico es una descripción realista de la respuesta verdadera e implica una prueba de la hipótesis $H_0: \beta_2 = 0$ contra $H_a: \beta_2 \neq 0$ en el modelo lineal $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ que se ajustó a los datos del Ejemplo 11.13. (Si $\beta_2 = 0$, el término cuadrático no aparecerá y el valor esperado de Y representará una recta en función de x.) El primer paso en la solución es calcular SSE y s^2:

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 11 - [.571 \quad .700 \quad .214]\begin{bmatrix} 5 \\ 7 \\ 13 \end{bmatrix}$$

$$= 11 - 10.537 = .463,$$

así que

$$s^2 = \frac{\text{SSE}}{n-3} = \frac{.463}{2} = .232 \quad \text{y} \quad s = .48.$$

(Observe que el modelo contiene tres parámetros y, por tanto, la SSE está basada en $n - 3 = 2$ grados de libertad.) El parámetro β_2 es una combinación lineal de β_0, β_1 y β_2 con $a_0 = 0$, $a_1 = 0$ y $a_2 = 1$. Para esta selección de $\mathbf{a}$, tenemos $\beta_2 = \mathbf{a}'\beta$ y $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = c_{22}$.

Los cálculos en el Ejemplo 11.13 dan $\hat{\beta}_2 = 3/14 \approx .214$ y $c_{22} = 1/14$. Por tanto, el estadístico de prueba apropiado se puede escribir como

$$t = \frac{\hat{\beta}_2 - 0}{s\sqrt{c_{22}}} = \frac{.214}{.48\sqrt{1/14}} = 1.67.$$

Si tomamos $\alpha = .05$, el valor de $t_{\alpha/2} = t_{.025}$ para 2 grados de libertad es 4.303 y la región de rechazo es

$$\text{rechazar si } |t| \geq 4.303.$$

Como el valor absoluto del valor calculado de t es menor que 4.303, no podemos rechazar la hipótesis nula de que $\beta_2 = 0$. No aceptamos $H_0: \beta_2 = 0$ porque necesitaríamos conocer la probabilidad de cometer un error tipo II —es decir, la probabilidad de aceptar erróneamente H_0 para un valor alternativo especificado de β_2— antes de que pudiéramos tomar una decisión estadísticamente sólida para aceptar. Debido a que la prueba es de dos colas, el valor $p = 2P(t > 1.67)$, donde t tiene una distribución t con 2 grados de libertad. Usando la Tabla 5, Apéndice 3, encontramos que $P(t > 1.67) > .10$. Por tanto, concluimos que el valor $p > .2$. Más precisamente, la aplicación *Student's t Probabilities and Quantiles* se puede usar para establecer que el valor $p = 2P(t > 1.67) = 2(.11843) = .23686$. A menos que estemos dispuestos a trabajar con un valor relativamente grande de α (al menos .23686), no podemos rechazar H_0. De nuevo observamos la correspondencia entre las conclusiones alcanzadas por el procedimiento de prueba formal (α fija) y la interpretación apropiada del nivel de significancia alcanzado.

Como un paso adicional en el análisis podríamos observar el ancho de un intervalo de confianza para β_2 para ver si es tan corto como para detectar una desviación desde cero que sea de importancia práctica. El intervalo de confianza de 95% resultante para β_2 es

$$\hat{\beta}_2 \pm t_{.025} S\sqrt{c_{22}}.$$

Sustituyendo, obtenemos

$$.214 \pm (4.303)(.48)\sqrt{1/14}, \quad \text{o} \quad .214 \pm .552.$$

Por consiguiente, el intervalo de confianza para β_2 es bastante ancho, lo cual sugiere que el experimentador necesita recolectar más información antes de tomar una decisión ■

EJEMPLO 11.16 Para los datos del Ejemplo 11.1 encuentre un intervalo de confianza de 90% para $E(Y)$ cuando $x = 1$.

Solución Para el modelo del Ejemplo 11.1,

$$E(Y) = \beta_0 + \beta_1 x = \mathbf{a}'\beta, \quad \text{con } \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 \\ x \end{bmatrix}.$$

El intervalo de confianza buscado está dado por

$$\mathbf{a}'\hat{\beta} \pm t_{\alpha/2} S\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

En el Ejemplo 11.12 determinamos que

$$\hat{\beta} = \begin{bmatrix} 1 \\ .7 \end{bmatrix} \quad \text{y} \quad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Como estamos interesados en $x = 1$,

$$\mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \mathbf{a}'\hat{\beta} = [1 \quad 1]\begin{bmatrix} 1 \\ .7 \end{bmatrix} = 1.7,$$

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = [1 \quad 1]\begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = .3.$$

En el Ejemplo 11.14 encontramos que s^2 es .367, o $s = .606$ para estos datos. El valor de $t_{.05}$ con $n - 2 = 3$ grados de libertad es 2.353 y el intervalo de confianza de 90% requerido para $E(Y)$ está dado por

$$1.7 \pm (2.353)(.606)\sqrt{.3}, \quad \text{o} \quad 1.7 \pm .781$$

Nuestra respuesta aquí es la misma que se obtuvo en el Ejemplo 11.6 sin el uso de matrices. ■

Ejercicios

11.71 Considere el modelo lineal general

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon,$$

donde $E(\varepsilon) = 0$ y $V(\varepsilon) = \sigma^2$. Observe que $\hat{\beta}_i = \mathbf{a}'\hat{\beta}$, donde el vector **a** está definido por

$$a_j = \begin{cases} 1, & \text{si } j = i, \\ 0, & \text{si } j \neq i. \end{cases}$$

Utilice este dato para verificar que $E(\hat{\beta}_i) = \beta_i$ y $V(\hat{\beta}_i) = c_{ii}\sigma^2$, donde c_{ii} es el elemento en la fila i y la columna i de $(\mathbf{X}'\mathbf{X})^{-1}$.

11.72 Consulte el Ejercicio 11.69.

a ¿Hay evidencia de un efecto cuadrático en la relación entre Y y x? (Pruebe H_0: $\beta_2 = 0$.) Use $\alpha = .10$.

b Encuentre un intervalo de confianza de 90% para β_2.

11.73 El experimentador que recolectó los datos del Ejercicio 11.68 afirma que el valor *mínimo* de $E(Y)$ se presenta en $x = 1$. Pruebe esta afirmación con un nivel de significancia de 5%. [*Sugerencia*: $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2$ tiene su mínimo en el punto x_0, que satisface la ecuación $\beta_1 + 2\beta_2 x_0 = 0$.]

11.74 Se realizó un experimento para investigar el efecto de cuatro factores: temperatura T_1, presión P, catalizador C y temperatura T_2, en la producción Y de una sustancia química.

a Los valores (o niveles) de los cuatro factores empleados en el experimento se muestran en la siguiente tabla. Si cada uno de los cuatro factores se codifica para generar las cuatro variables x_1, x_2, x_3 y x_4, respectivamente, indique la transformación que relaciona cada variable codificada con su original correspondiente.

T_1	x_1	P	x_2	C	x_3	T_2	x_4
50	−1	10	−1	1	−1	100	−1
70	1	20	1	2	1	200	1

b Ajuste el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

a la siguiente tabla de datos.

				x_4			
				+1		**−1**	
				x_3		x_3	
				−1	**1**	**−1**	**1**
x_1	**−1**	x_2	**−1**	22.2	24.5	24.4	25.9
			1	19.4	24.1	25.2	28.4
			−1	22.1	19.6	23.5	16.5
	+1	x_2	**1**	14.2	12.7	19.3	16.0

c ¿Los datos presentan suficiente evidencia para indicar que T_1 contribuye con información para el cálculo de Y?, ¿y P?, ¿y C?, ¿y T_2? (Pruebe las hipótesis, respectivamente, de que $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$ y $\beta_4 = 0$.) Dé límites para el valor p asociado con cada prueba. ¿Qué concluiría usted si utiliza $\alpha = .01$ en cada caso?

11.75 Consulte el Ejercicio 11.74. Encuentre un intervalo de confianza de 90% para la producción esperada, dado que $T_1 = 50$, $P = 20$, $C = 1$ y $T_2 = 200$.

11.76 Los siguientes resultados se obtuvieron de un análisis de datos obtenido de un estudio para evaluar la relación entre porcentaje de aumento en rendimiento (Y) y saturación de base (x_1, libras/acre), saturación de fosfato (x_2, BEC%) y pH del suelo (x_3). Aleatoriamente se analizaron quince respuestas del estudio. A continuación aparece la ecuación de mínimos cuadrados y otra información útil.

$$\hat{y} = 38.83 - 0.0092x_1 - 0.92x_2 + 11.56x_3, \qquad S_{yy} = 10965.46, \qquad \text{SSE} = 1107.01,$$

$$10^4(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 15\,1401.8 & 2.6 & 100.5 & -28\,082.9 \\ 2.6 & 1.0 & 0.0 & 0.4 \\ 100.5 & 0.0 & 8.1 & 5.2 \\ -28\,082.9 & 0.4 & 5.2 & 6038.2 \end{bmatrix}.$$

a ¿Hay suficiente evidencia de que, con todas las variables independientes en el modelo, $\beta_2 < 0$? Pruebe en el nivel de significancia $\alpha = .05$.

b Proporcione un intervalo de confianza de 95% para el porcentaje medio de aumento en rendimiento si $x_1 = 914$, $x_2 = 65$ y $x_3 = 6$.

11.13 Predicción de un valor particular de *Y* mediante regresión múltiple

En la Sección 11.7 consideramos predecir un valor real observado de Y en la regresión lineal *simple*, haciendo la variable independiente individual $x = x^*$. La solución estuvo basada fuertemente en las propiedades de

$$\text{error} = Y^* - \widehat{Y^*},$$

donde se observó que $\widehat{Y^*} = \hat{\beta}_0 + \beta_1 x^*$ era un pronosticador del valor real de Y y también un estimador para $E(Y)$. El mismo método se utilizará en esta sección para obtener la solución correspondiente en el caso de una regresión lineal *múltiple*. Suponga que hemos ajustado un modelo de regresión lineal múltiple

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

y que estamos interesados en predecir el valor de Y^* cuando $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_k = x_k^*$. Predecimos el valor de Y^* con

$$\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \hat{\beta}_2 x_2^* + \cdots + \hat{\beta}_k x_k^* = \mathbf{a}'\hat{\beta},$$

donde

$$\mathbf{a} = \begin{bmatrix} 1 \\ x_1^* \\ x_2^* \\ \vdots \\ x_k^* \end{bmatrix}.$$

Al igual que en la Sección 11.7 nos concentramos en la diferencia entre la variable Y^* y el valor predicho:

$$\text{error} = Y^* - \widehat{Y^*}.$$

Como Y^* y $\widehat{Y^*}$ están distribuidas normalmente, el error está distribuido normalmente; si aplicamos el Teorema 5.12 y los resultados de la Sección 11.11, encontramos que

$$E(\text{error}) = 0 \quad \text{y} \quad V(\text{error}) = \sigma^2[1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]$$

y que

$$Z = \frac{Y^* - \widehat{Y^*}}{\sigma\sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}x$$

tiene una distribución normal estándar. Además, si S es sustituido por σ, se puede demostrar que

$$T = \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

tiene una distribución t de Student con $[n - (k + 1)]$ grados de libertad.

Si procedemos como en la Sección 11.7, obtenemos el siguiente intervalo de predicción $100(1 - \alpha)\%$ para Y.

Un intervalo de predicción $100(1 - \alpha)\%$ para Y cuando $x_1 = x_1^*$, $x_2 = x_2^*, \ldots, x_k = x_k^*$

$$\mathbf{a}'\hat{\beta} \pm t_{\alpha/2}S\sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}},$$

donde $\mathbf{a}' = [1, x_1^*, x_2^*, \ldots, x_k^*]$.

EJEMPLO **11.17** Suponga que el experimento que generó los datos del Ejemplo 11.12 se va a ejecutar de nuevo con $x = 2$. Prediga el valor particular de Y con $1 - \alpha = .90$.

Solución En el Ejemplo 11.12 determinamos que

$$\hat{\beta} = \begin{bmatrix} 1 \\ .7 \end{bmatrix} \quad \text{y} \quad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}.$$

Debido a que estamos interesados en $x = 2$, el intervalo de predicción deseado está dado por

$$\mathbf{a}'\hat{\beta} \pm t_{\alpha/2}S\sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

con

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \mathbf{a}'\hat{\beta} = [1 \quad 2]\begin{bmatrix} 1 \\ .7 \end{bmatrix} = 2.4,$$

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = [1 \quad 2]\begin{bmatrix} 1/5 & 0 \\ 0 & 1/10 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = .6.$$

Al igual que antes, $s = .606$ para estos datos y el valor de $t_{.05}$ con $n - 2 = 3$ grados de libertad es 2.353. El intervalo de predicción de 90% para una observación futura en Y cuando $x = 2$ es, por tanto,

$$2.4 \pm (2.353)(.606)\sqrt{1 + .6}, \quad \text{o} \quad 2.4 \pm 1.804.$$

Observe la concordancia del resultado con la respuesta proporcionada en el Ejemplo 11.7 en donde empleamos álgebra ordinaria en lugar del método matricial en la solución. ■

Ejercicios

11.77 Consulte el Ejercicio 11.76. Obtenga un intervalo de predicción de 95% para el porcentaje de aumento en rendimiento en un campo con saturación de base = 914 libras/acre, saturación de fosfato = 65% y pH del suelo = 6.

11.78 Consulte el Ejercicio 11.69. Encuentre un intervalo de predicción de 98% para las ventas del Lexus en 2004. Use el modelo cuadrático.

11.79 Consulte los Ejercicios 11.74 y 11.75. Encuentre un intervalo de predicción de 90% para Y si $T_1 = 50$, $P = 20$, $C = 1$ y $T_2 = 200$.

11.14 Una prueba para H_0: $\beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0$

Al buscar un estadístico de prueba razonablemente atractivo para probar una hipótesis relacionada con un conjunto de parámetros del modelo lineal, llegamos a la consideración de la suma de cuadrados de las desviaciones SSE. Suponga, por ejemplo, que debemos ajustar un modelo que comprende sólo un subconjunto de las variables independientes en consideración, esto es, ajustar un modelo *reducido* de la forma

$$\text{modelo R}: Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_g x_g + \varepsilon$$

a los datos y que luego debemos calcular la suma de cuadrados de las desviaciones entre los valores observados y pronosticados de Y, SSE_R. Habiendo hecho esto, podríamos ajustar el modelo lineal con *todas* las variables independientes propuestas presentes (el modelo *completo*):

$$\text{modelo C}: Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_g x_g + \beta_{g+1} x_{g+1} + \cdots + \beta_k x_k + \varepsilon$$

y determinar la suma de cuadrados de las desviaciones para este modelo, SSE_C. Observe que el modelo completo contiene todos los términos del modelo reducido, modelo R, más los términos extra $x_{g+1}, x_{g+2}, \ldots, x_k$ (observe que $k > g$). Si $x_{g+1}, x_{g+2}, \ldots, x_k$ contribuye con una cantidad importante de información para la predicción de Y que no esté contenida en las variables $x_1, x_2, \ldots, x_g$ (esto es, al menos uno de los parámetros $\beta_{g+1}, \beta_{g+2}, \ldots, \beta_k$ difiere de cero), ¿cuál sería la relación entre SSE_R y SSE_C? Intuitivamente vemos que si $x_{g+1}, x_{g+2}, \ldots, x_k$ son variables que aportan información importante, el modelo C, el modelo *completo* debe predecir con un error de predicción *menor* que el modelo R. Esto es, SSE_C debe ser menor que SSE_R. Cuanto mayor sea la diferencia $(\text{SSE}_R - \text{SSE}_C)$, más fuerte será la evidencia para apoyar la hipótesis alternativa de que $x_{g+1}, x_{g+2}, \ldots, x_k$ contribuye con información para la predicción de Y y para rechazar la hipótesis nula

$$H_0: \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0.$$

La disminución en la suma de cuadrados de las desviaciones ($\text{SSE}_R - \text{SSE}_C$) recibe el nombre de *suma de cuadrados asociada con las variables* $x_{g+1}, x_{g+2}, \ldots, x_k$, *ajustada para las variables* $x_1, x_2, x_3, \ldots, x_g$.

Indicamos que valores grandes de ($\text{SSE}_R - \text{SSE}_C$) nos llevarían a rechazar la hipótesis

$$H_0: \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0.$$

¿Qué tan grande es "grande"? Crearemos un estadístico de prueba que es una función de ($\text{SSE}_R - \text{SSE}_C$) para el cual conocemos la distribución cuando H_0 es verdadera.

Para obtener este estadístico de prueba, *supongamos* que la hipótesis nula es verdadera y luego examinemos las cantidades que hemos calculado. En particular, observe que

$$\text{SSE}_R = \text{SSE}_C + (\text{SSE}_R - \text{SSE}_C).$$

En otras palabras, como indicamos en la Figura 11.8, hemos dividido SSE_R en dos partes: SSE_C y la diferencia ($\text{SSE}_R - \text{SSE}_C$). Aunque omitimos la prueba, si H_0 es verdadera, entonces

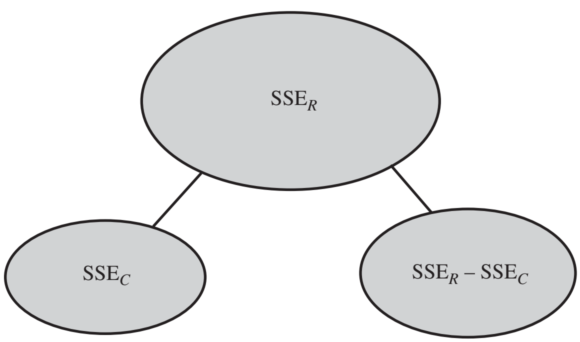
$$\chi_3^2 = \frac{\text{SSE}_R}{\sigma^2},$$
$$\chi_2^2 = \frac{\text{SSE}_C}{\sigma^2},$$
$$\chi_1^2 = \frac{\text{SSE}_R - \text{SSE}_C}{\sigma^2}$$

tiene distribuciones de probabilidad χ^2 en muestreo repetitivo, con $(n - [g + 1])$, $(n - [k + 1])$, y $(k - g)$ grados de libertad, respectivamente. Además, se puede demostrar que χ_2^2 y χ_1^2 son estadísticamente independientes.

La definición de una variable aleatoria con una distribución F se da en la Definición 7.3. Considere la razón

$$F = \frac{\chi_1^2/(k-g)}{\chi_2^2/(n-[k+1])} = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{(\text{SSE}_C)/(n-[k+1])}.$$

FIGURA **11.8**
División de SSE_R



Si $H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0$ es verdadera, entonces F tiene una distribución F con $v_1 = k - g$ grados de libertad en el numerador y $v_2 = n - (k + 1)$ grados de libertad en el denominador. Ya hemos dicho que valores grandes de $(\text{SSE}_R - \text{SSE}_C)$ nos llevan a rechazar la hipótesis nula. Entonces, valores grandes de F favorecen el rechazo de H_0; si deseamos una prueba con una probabilidad de error tipo I igual a α, se deduce que

$$F > F_\alpha$$

es la región de rechazo apropiada. (Véase la Tabla 7, Apéndice 3.)

EJEMPLO **11.18** ¿Los datos del Ejemplo 11.13 proporcionan suficiente evidencia para indicar que el modelo de segundo orden

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

aporta información para la predicción de Y? Esto es, pruebe la hipótesis $H_0 : \beta_1 = \beta_2 = 0$ contra la hipótesis alternativa H_a: al menos uno de los parámetros β_1, β_2 difiere de 0. Use $\alpha =$ .05. Obtenga límites para el nivel de significancia alcanzado.

Solución Para el modelo completo, determinamos en el Ejemplo 11.15 que $\text{SSE}_C = .463$. Como deseamos probar $H_0 : \beta_1 = \beta_2 = 0$, el modelo reducido apropiado es

$$Y = \beta_0 + \varepsilon$$

para el cual

$$\mathbf{Y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{bmatrix} \quad \text{y} \quad \mathbf{X} = \begin{array}{c} x_0 \\ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{array}.$$

Porque $\mathbf{X}'\mathbf{X} = 5$, $(\mathbf{X}'\mathbf{X})^{-1} = 1/5$ y $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (1/5)\sum_{i=1}^{5} y_i = \overline{y} = 5/5 = 1$
Entonces,

$$\begin{aligned}
\text{SSE}_R &= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} \\
&= \sum_{i=1}^{5} y_i^2 - \overline{y}\left(\sum_{i=1}^{n} y_i\right) = \sum_{i=1}^{5} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{5} y_i\right)^2 \\
&= 11 - (1/5)(5)^2 = 11 - 5 = 6.
\end{aligned}$$

En este ejemplo, el número de variables independientes del modelo completo es $k = 2$, y el número de variables independientes del modelo reducido es $g = 0$. Por tanto,

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k - g)}{(\text{SSE}_C)/(n - [k + 1])} = \frac{(6 - .463)/(2 - 0)}{.463/(5 - 3)} = 11.959.$$

El valor F tabulado para $\alpha = .05$ con $v_1 = k - g = 2$ grados de libertad en el numerador y $v_2 = n - (k + 1) = 2$ grados de libertad en el denominador es 19.00. Por tanto, el valor observado del estadístico de prueba no cae en la región de rechazo y concluiríamos que en el nivel de $\alpha = .05$ no hay suficiente evidencia para apoyar la afirmación de que β_1 o β_2 difiera de cero. Debido a que la forma apropiada de la región de rechazo es $F > F_\alpha$, el valor p está dado por $P(F > 11.959)$ cuando F está basada en 2 grados de libertad en el numerador y 2 en el denominador. Usando la Tabla 7, Apéndice 3, usted puede ver que $.05 <$ valor $p < .10$. Además, la aplicación *F-Ratio Probabilities and Quantiles* da $P(F > 11.959) = .07717$. En consecuencia, si escogemos $\alpha = .05$ (de acuerdo con la exposición previa), no hay suficiente evidencia para apoyar la afirmación de que β_1 o β_2 difiera de cero. No obstante, si se selecciona cualquier valor α igual o mayor que .0772 podríamos decir que ya sea $\beta_1 \neq 0$ o que $\beta_2 \neq 0$. Obsérvese que el pequeño esfuerzo adicional requerido para determinar el valor p aporta una considerable cantidad de información adicional. ■

Considere la situación donde hemos ajustado un modelo con k variables independientes y deseamos probar la hipótesis nula

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

de que *ninguna* de las variables independientes del modelo contribuye con información importante para la predicción de Y. Esto es exactamente lo que se hizo en el Ejemplo 11.18. Un examen de la solución de ese ejemplo convencerá al lector de que el modelo *reducido* apropiado es de la forma

$$Y = \beta_0 + \varepsilon.$$

Este modelo reducido contiene $g = 0$ variables independientes y es tal que $\text{SSE}_R = S_{yy}$ (vea el Ejemplo 11.18). Por tanto, una prueba para

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

puede estar basada en el estadístico

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{(\text{SSE}_C)/(n-[k+1])} = \frac{(S_{yy} - \text{SSE}_C)/k}{(\text{SSE}_C)/(n-[k+1])},$$

que tiene una distribución F con $v_1 = k$ y $v_2 = n - (k + 1)$ grados de libertad en el numerador y el denominador, respectivamente.

¿Qué proporción de la variación en los valores observados de la variable de respuesta, Y, es explicada por todo el conjunto de variables independientes $x_1, x_2, \ldots, x_k$? La respuesta está dada por el *coeficiente múltiple de determinación* R^2, donde

$$R^2 = \frac{S_{yy} - \text{SSE}_C}{S_{yy}}.$$

Al igual que con el coeficiente sencillo de determinación r^2, el denominador de R^2 cuantifica la variación en los valores y, y el numerador cuantifica la cantidad de variación en las y que está explicada por el conjunto completo de variables independientes $x_1, x_2, \ldots, x_k$. En el Ejercicio 11.84(a), usted demostrará que el estadístico F para probar

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

se puede calcular usando R^2 mediante la fórmula

$$F = \frac{n-(k+1)}{k}\left(\frac{R^2}{1-R^2}\right).$$

Al igual que antes, este estadístico tiene una distribución F con $v_1 = k$ y $v_2 = n - (k + 1)$ grados de libertad en el numerador y en el denominador, respectivamente.

Otra aplicación del método general para comparar modelos completos y reducidos se da en el siguiente ejemplo.

EJEMPLO 11.19 Se desea relacionar la resistencia a la abrasión del caucho (Y) con la cantidad de relleno de sílice x_1' y la cantidad de agente de acoplamiento x_2'. Se agregan fibras de sílice de partículas finas al caucho para aumentar su fuerza y resistencia a la abrasión. El agente de acoplamiento enlaza químicamente el relleno a las cadenas de polímeros del caucho y en esa forma aumenta la eficiencia del relleno. La unidad de medida para x_1' y x_2' es partes por 100 partes de caucho, lo cual se denota como phr. Para mayor simplicidad en los cálculos, las cantidades reales de relleno de sílice y agente de acoplamiento se cambian de escala mediante las ecuaciones

$$x_1 = \frac{x_1' - 50}{6.7} \quad \text{y} \quad x_2 = \frac{x_2' - 4}{2}.$$

(Este cambio de escala de las variables independientes no afecta el análisis ni las conclusiones, pero simplifica los cálculos.)

Los datos[11] se proporcionan en la Tabla 11.6. Observe que se emplean cinco niveles de x_1 y de x_2, con el punto ($x_1 = 0$, $x_2 = 0$) repetido tres veces. Ajustemos el modelo de segundo orden

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon$$

a estos datos. Este modelo representa una superficie cónica sobre el plano (x_1, x_2). Ajuste el modelo de segundo orden y pruebe $H_0: \beta_3 = \beta_4 = \beta_5 = 0$. (Estamos probando que la superficie es en realidad un plano contra la alternativa de que es una superficie cónica.) Indique los límites para el nivel de significancia alcanzado e indique la conclusión apropiada si escogemos $\alpha = .05$.

Solución Primero usaremos ecuaciones matriciales para ajustar el modelo completo, como ya indicamos antes. (Con modelos de este tamaño, es mejor usar una computadora para hacer los cálculos.) Para los datos de la Tabla 11.6 tenemos

Tabla 11.6 Datos para el Ejemplo 11.19

y	x_1	x_2
83	1	−1
113	1	1
92	−1	1
82	−1	−1
100	0	0
96	0	0
98	0	0
95	0	1.5
80	0	−1.5
100	1.5	0
92	−1.5	0

11. *Fuente*: Ronald Suich y G. C. Derringer, *Technometrics* 19(2) (1977): 214.

$$
\mathbf{Y} = \begin{bmatrix} 83 \\ 113 \\ 92 \\ 82 \\ 100 \\ 96 \\ 98 \\ 95 \\ 80 \\ 100 \\ 92 \end{bmatrix}, \qquad
\begin{array}{cc}
 & \begin{array}{cccccc} & x_1 & x_2 & x_1^2 & x_2^2 & x_1x_2 \end{array} \\
\mathbf{X} = & \begin{bmatrix}
1 & 1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 & 1 & -1 \\
1 & -1 & -1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1.5 & 0 & 2.25 & 0 \\
1 & 0 & -1.5 & 0 & 2.25 & 0 \\
1 & 1.5 & 0 & 2.25 & 0 & 0 \\
1 & -1.5 & 0 & 2.25 & 0 & 0
\end{bmatrix},
\end{array}
$$

$$
(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix}
.33 & 0 & 0 & -.15 & -.15 & 0 \\
0 & 0.12 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.12 & 0 & 0 & 0 \\
-.15 & 0 & 0 & .15 & .05 & 0 \\
-.15 & 0 & 0 & .05 & .15 & 0 \\
0 & 0 & 0 & 0 & 0 & .25
\end{bmatrix}.
$$

Estas matrices dan como resultado

$$
\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 98.00 \\ 4.00 \\ 7.35 \\ -.88 \\ -4.66 \\ 5.00 \end{bmatrix},
$$

o el modelo ajustado de segundo orden,

$$
\hat{y} = 98.00 + 4.00x + 7.35x_2 - .88x_1^2 - 4.66x_2^2 + 5.00x_1x_2.
$$

En el caso de este modelo, $\text{SSE}_C = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 77.948$.

Para probar la hipótesis de interés ($H_0: \beta_3 = \beta_4 = \beta_5 = 0$), debemos ajustar el modelo reducido

$$
Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.
$$

Si borramos las columnas para x_1^2, x_2^2, y x_1x_2 en la matriz $\mathbf{X}$, tenemos

$$
\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 93.73 \\ 4.00 \\ 7.35 \end{bmatrix},
$$

y el modelo plano ajustado es

$$
\hat{y} = 93.73 + 4.00x_1 + 7.35x_2.
$$

(Observe que no podemos simplemente igualar $\hat{\beta}_3$, $\hat{\beta}_4$, y $\hat{\beta}_5$ a cero para producir el modelo ajustado en el caso reducido.) Para el modelo reducido, $\text{SSE}_R = 326.623$.

Ahora probamos la hipótesis $H_0: \beta_3 = \beta_4 = \beta_5 = 0$ al calcular F (Observe que $k = 5$, $g = 2$ y $n = 11$):

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{\text{SSE}_C/[n-(k+1)]} = \frac{(326.623 - 77.948)/3}{77.948/5} = 5.32.$$

Como este estadístico se basa en $v_1 = (k - g) = 3$ grados de libertad en el numerador y $v_2 = n - (k + 1) = 5$ grados de libertad en el denominador, el valor p está dado por $P(F > 5.32)$. Entonces, usando la Tabla 7, Apéndice 3, $.05 <$ valor $p < .10$. La aplicación *F-Ratio Probabilities and Quantiles* proporciona el valor p exacto $= P(F > 5.32) = .05155$. Si escogemos $\alpha = .05$, no hay suficiente evidencia para apoyar la afirmación de que el modelo de segundo orden ajusta los datos considerablemente mejor que el modelo plano. ¿El valor p exacto $= .05155$ es lo suficientemente pequeño para convencerlo de que el modelo de segundo orden ajusta mejor que el modelo plano? Sólo usted puede contestar esa pregunta. Observe que hemos probado si el *grupo* de variables x_1^2, x_2^2, x_1x_2 contribuyó a un considerablemente mejor ajuste del modelo a los datos. ■

Ejercicios

11.80 Consulte el Ejercicio 11.31. Conteste la pregunta sobre el aumento en la corriente pico al construir una prueba F.

11.81 En el Ejercicio 11.80 usted empleó una prueba F para comprobar la misma hipótesis que se probó en el Ejercicio 11.31 por medio de una prueba t. Considere el caso general de regresión lineal simple y los estadísticos F y t que se puedan usar para poner en práctica la prueba de $H_0: \beta_1 = 0$ contra $H_a: \beta_1 \neq 0$. Demuestre que en general $F = t^2$. Compare el valor de F obtenido en el Ejercicio 11.80 con el correspondiente valor de t obtenido en el Ejercicio 11.31.

11.82 Consulte el Ejercicio 11.76 donde obtuvimos la siguiente información al ajustar un modelo de regresión simple a 15 respuestas;

$$\hat{y} = 38.83 - 0.0092x_1 - 0.92x_2 + 11.56x_3, \qquad S_{yy} = 10\,965.46, \qquad \text{SSE} = 1107.01.$$

a ¿Hay suficiente evidencia para concluir que al menos una de las variables independientes contribuye con información importante para el pronóstico de Y?

b Calcule el valor del coeficiente de determinación múltiple. Interprete el valor de R^2.

11.83 Consulte los Ejercicios 11.76 y 11.82. ¿La inclusión de las variables de saturación de fosfato x_2 y pH x_3 contribuye a un ajuste significativamente mejor del modelo a los datos? El modelo de regresión lineal reducido $Y = \beta_0 + \beta_1 x_1 + \varepsilon$ se ajustó y observamos $\text{SSE}_R = 5470.07$.

a Lleve a cabo la prueba apropiada de hipótesis en el nivel de significancia $\alpha = .05$.

b ¿Cuál es el mínimo valor de SSE_R que le hubiera permitido concluir que al menos una de las variables (saturación de fosfato y/o pH) contribuyó a un mejor ajuste del modelo a los datos?

11.84 Hemos ajustado un modelo con k variables independientes y deseamos probar la hipótesis nula $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$.

a Demuestre que el estadístico de prueba apropiado con distribución F se puede expresar como

$$F = \frac{n - (k + 1)}{k}\left(\frac{R^2}{1 - R^2}\right).$$

b Si $k = 1$, ¿cómo se compara el valor de F del inciso a con la expresión para el estadístico T deducido en el Ejercicio 11.55?

11.85 Los datos computarizados de un agente de bienes raíces contienen la lista de precios de venta Y (en miles de dólares), el área de vivienda x_1 (en cientos de pies cuadrados), el número de pisos x_2, el número de alcobas x_3 y el número de baños x_4 para condominios recién registrados. El modelo de regresión múltiple $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$ se ajustó a los datos obtenidos al seleccionar de manera aleatoria 15 departamentos actualmente en el mercado.

a Si $R^2 = .942$, ¿hay suficiente evidencia de que al menos una de las variables independientes contribuya con información importante para el pronóstico del precio de venta?

b Si $S_{yy} = 16\,382.2$, ¿cuál es SSE?

11.86 Consulte el Ejercicio 11.85. Un corredor de bienes raíces sospecha que la superficie en pies cuadrados, x_1, podría ser la variable de pronóstico más importante y que las otras variables se pueden eliminar del modelo sin mucha pérdida de información en la predicción. El modelo de regresión lineal simple para precio de venta contra superficie en pies cuadrados se ajustó a los 15 datos que se emplearon en el Ejercicio 11.85 y el corredor de bienes raíces observó que SSE $= 1553$. Las variables independientes adicionales usadas para ajustar el modelo en el Ejercicio 11.85 ¿pueden cancelarse del modelo sin perder información de predicción? Pruebe en el nivel de significancia $\alpha = .05$.

11.87 ¿El valor grande de R^2 siempre implica que al menos una de las variables independientes deba ser retenida en el modelo de regresión? ¿Un valor pequeño de R^2 siempre indica que ninguna de las variables independientes son útiles para la predicción de la respuesta?

a Suponga que un modelo con $k = 4$ variables independientes se ajusta usando $n = 7$ datos y que $R^2 = .9$. ¿Cuántos grados de libertad en el numerador y en el denominador están asociados con el estadístico F para probar $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$? Use el resultado del Ejercicio 11.84(a) para calcular el valor del estadístico F apropiado. ¿Puede H_0 ser rechazada en el nivel de significancia de $\alpha = .10$?

b Consulte el inciso a. ¿Qué observa usted acerca de los tamaños relativos de n y k? ¿Qué impacto tiene esto en el valor de F?

c Un modelo con $k = 3$ variables independientes se ajusta a $n = 44$ datos que resultan en $R^2 = .15$. ¿Cuántos grados de libertad en el numerador y en el denominador están asociados con el estadístico F para probar $H_0: \beta_1 = \beta_2 = \beta_3 = 0$? Use el resultado del Ejercicio 11.84(a) para calcular el valor del estadístico F apropiado. ¿Puede H_0 ser rechazada en el nivel de significancia de $\alpha = .10$?

d Consulte el inciso c. ¿Qué observa acerca de los tamaños relativos de n y k? ¿Qué impacto tiene esto en el valor de F?

11.88 La publicidad en televisión, en el caso ideal, estaría destinada exactamente a la audiencia que ve los anuncios. Se realizó un estudio para determinar el tiempo que las personas pasan viendo TV durante las horas de mayor audiencia al oscurecer. Se observaron veinte personas durante un periodo de una semana y de cada una se registró el tiempo promedio que pasaban viendo TV al oscurecer, llamado Y. También se registraron otros cuatro datos para cada individuo: x_1 = edad, x_2 = grado de escolaridad, x_3 = ingreso disponible y x_4 = cociente de inteligencia. Considere los tres modelos dados a continuación:

Modelo I: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$
Modelo II: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
Modelo III: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

¿Las siguientes frases son verdaderas o falsas?

a Si se ajusta el modelo I, la estimación para σ^2 está basada en 16 grados de libertad.

b Si se ajusta el modelo II, podemos realizar la prueba t para determinar si x_2 contribuye a un mejor ajuste del modelo a los datos.

c Si los modelos I y II se ajustan, entonces $\text{SSE}_{\text{I}} \leq \text{SSE}_{\text{II}}$.

d Si los modelos I y II se ajustan, entonces $\hat{\sigma}^2_{\text{I}} \leq \hat{\sigma}^2_{\text{II}}$.

e El modelo II es una reducción del modelo I.

f Los modelos I y III se pueden comparar usando la técnica de modelo completo/reducido presentado en la Sección 11.14.

11.89 Consulte los tres modelos dados en el Ejercicio 11.88. Denote con R^2_{I}, R^2_{II} y R^2_{III} los coeficientes de determinación para los modelos I, II y III. ¿Las expresiones siguientes son verdaderas o falsas?

a $R^2_{\text{I}} \geq R^2_{\text{II}}$.

b $R^2_{\text{I}} \geq R^2_{\text{III}}$.

c $R^2_{\text{II}} \leq R^2_{\text{III}}$.

11.90 Consulte el Ejercicio 11.69.

a Para el modelo cuadrático realice una prueba F de $H_0: \beta_2 = 0$, usando $\alpha = .05$. Compare el resultado con la solución de la prueba en el Ejercicio 11.72.

b Pruebe $H_0: \beta_1 = \beta_2 = 0$ con un nivel de significancia de 5%.

11.91 Consulte el Ejercicio 11.74. Pruebe la hipótesis con un nivel de significancia de 5% de que ni T_1 ni T_2 afectan la producción.

11.92 Las empresas de energía eléctrica, que deben planear la operación y expansión de generación de electricidad, están muy interesadas en pronosticar la demanda de clientes tanto a corto como a largo plazos. Se realizó un estudio a corto plazo para investigar el efecto de la temperatura x_1 media diaria de cada mes y el costo por kilowatt-hora, x_2 en el consumo medio diario (en kWh) por familia. Los directores de la compañía esperaban que la demanda de electricidad subiera en climas fríos (debido a la calefacción empleada), bajara cuando la temperatura fuera moderada y volviera a subir la demanda cuando la temperatura aumentara y fuera necesario usar el aire acondicionado. Esperaban que la demanda se redujera a medida que subiera el costo por kilowatt-hora, reflejando mayor atención a la conservación. Se obtuvieron datos durante 2 años, periodo durante el cual el costo por kilowatt-hora x_2 aumentó debido a los crecientes costos del combustible. Los directores de la compañía ajustaron el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2 + \varepsilon$$

a los datos de la siguiente tabla y obtuvieron $\hat{y} = 325.606 - 11.383x_1 + .113x_1^2 - 21.699x_2 + .873x_1x_2 - .009x_1^2x_2$ con SSE $= 152.1776$.

Precio por kWh (x_2)		Consumo medio diario (kWh) por familia					
8¢	Temperatura media diaria en °F (x_1)	31	34	39	42	47	56
	Consumo medio diario (y)	55	49	46	47	40	43
10¢	Temperatura media diaria en °F (x_1)	32	36	39	42	48	56
	Consumo medio diario (y)	50	44	42	42	38	40
8¢	Temperatura media diaria en °F (x_1)	62	66	68	71	75	78
	Consumo medio diario (y)	41	46	44	51	62	73
10¢	Temperatura media diaria en °F (x_1)	62	66	68	72	75	79
	Consumo medio diario (y)	39	44	40	44	50	55

Cuando el modelo $Y = \beta_0 - \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon$ se ajustó, la ecuación de predicción fue $\hat{y} = 130.009$ $- 3.302x_1 + .033x_1^2$ con SSE = 465.134. Pruebe si los términos que contienen $x_2(x_2, x_1x_2, x_1^2x_2)$ contribuyen a un ajuste considerablemente mejor a los datos. Obtenga límites para el nivel de significancia alcanzado.

11.93 Consulte el Ejemplo 11.19. Usando el modelo reducido construya un intervalo de confianza de 95% para la resistencia esperada del caucho a la abrasión cuando $x_1 = 1$ y $x_2 = -1$.

11.94 Consulte el Ejemplo 11.19. Construya pruebas individuales de las tres hipótesis $H_0: \beta_3 = 0$, $H_0: \beta_4 = 0$ y $H_0: \beta_5 = 0$. Use un nivel de significancia de 1% en cada prueba. (Si han de realizarse pruebas múltiples en el mismo conjunto de datos, es conveniente usar un nivel α muy pequeño en cada prueba.)

11.15 Resumen y conclusiones

En este capítulo hemos empleado el método de mínimos cuadrados para ajustar un modelo lineal a una respuesta experimental. Supusimos que el valor esperado de Y es una función de un conjunto de variables $x_1, x_2, \ldots, x_k$, donde la función es lineal en un conjunto de parámetros desconocidos. Usamos la expresión

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

para denotar un modelo estadístico lineal.

Los problemas inferenciales asociados con el modelo estadístico lineal incluyen estimación y pruebas de hipótesis que se relacionan con los parámetros de modelo $\beta_0, \beta_1, \ldots, \beta_k$ y, lo que es más importante, el cálculo de $E(Y)$ que es la respuesta esperada para un ajuste particular y la predicción de algún valor futuro de Y. Experimentos para los cuales es apropiada la teoría de mínimos cuadrados incluyen experimentos controlados y aquellos en los que $x_1, x_2, \ldots, x_k$ son valores observados de variables aleatorias.

¿Por qué usar el método de mínimos cuadrados para ajustar un modelo lineal a un conjunto de datos? Donde las suposiciones acerca de los errores aleatorios ε se cumplen [normalidad, independencia, $V(\varepsilon) = \sigma^2$ para todos los valores de $x_1, x_2, \ldots, x_k$], se puede demostrar que el procedimiento de mínimos cuadrados da los mejores estimadores insesgados *lineales* para $\beta_0, \beta_1, \ldots, \beta_k$. Esto es, si calculamos los parámetros $\beta_0, \beta_1, \ldots, \beta_k$ usando funciones lineales de $y_1, y_2, \ldots, y_k$, los estimadores de mínimos cuadrados tienen varianza mínima. Algunos otros estimadores no lineales para los parámetros pueden tener una menor varianza que los estimadores de mínimos cuadrados, pero, si existen estos estimadores, no se conocen en este momento. De nueva cuenta, ¿por qué usar estimadores de mínimos cuadrados? Son fáciles de emplear y sabemos que tienen buenas propiedades para numerosas situaciones.

Como puede imaginarse, la metodología presentada en este capítulo tiene uso generalizado en finanzas y en todas las ciencias para explorar la relación entre una respuesta y un conjunto de variables independientes. La estimación de $E(Y)$ o la predicción de Y suele ser el objetivo experimental.

Libros de texto enteros están dedicados al tema de regresiones. Nuestro propósito ha sido introducir muchas de las consideraciones teóricas asociadas con regresión lineal simple y múltiple. Aunque el método de mínimos cuadrados se puede usar para calcular parámetros de modelo en situaciones generales, las técnicas formales de hacer inferencia que hemos presentado (con base en distribuciones t y F) son válidas sólo dadas las suposiciones adicionales que presentamos. Las suposiciones clave incluyen que los términos de error del modelo están distribuidos normalmente y que la varianza de los términos de error no depende del valor de ninguna(s) variable(s) independiente(s). En aplicaciones prácticas estas suposiciones pueden no ser válidas. Generalmente las evaluaciones de la validez de suposiciones de un modelo están basadas en análisis de los *residuos*, es decir, las diferencias entre valores observados y pronosticados (usando el modelo) de la variable de respuesta. El examen de los residuos, incluyendo gráficas de los residuos contra la(s) variable(s) independiente(s) y gráficas de los residuos contra sus valores esperados teóricos normales, permite evaluar si las suposiciones son razonables para un conjunto de datos en particular. Los datos con residuos anormalmente grandes pueden ser los *resultados atípicos* que indican que algo anduvo mal cuando se hizo la observación correspondiente. Algunos datos individuales pueden tener un impacto anormalmente grande en el modelo de regresión ajustado, en el sentido de que el modelo ajustado con estos datos incluidos difiere de manera considerable del modelo ajustado con ellos excluidos (estos datos suelen recibir el nombre de *puntos de alta influencia*; vea Ejercicio 11.108). Un modelo de regresión podría sufrir de *falta de ajuste*, lo cual indica que el modelo seleccionado no es adecuado para modelar la respuesta. En tales casos podría ser necesario ajustar un modelo más complicado para obtener precisión suficiente de predicción. Una consideración importante en los modelos de regresión múltiple es la de *multicolinealidad* en la que algunas de las variables independientes del modelo están correlacionadas de modo estrecho entre ellas. No podemos hacer justicia a estos temas en un solo capítulo introductorio sobre regresión lineal y múltiple. Nos hemos concentrado en el concepto general de mínimos cuadrados como un método para estimar parámetros de modelo y hemos dado las bases teóricas para el análisis basado en la teoría normal clásica. Los otros temas descritos en esta sección se tratan en las referencias adicionales.

Bibliografía y lecturas adicionales

Draper, N. R., and H. Smith. 1998. *Applied Regression Analysis*, 3d ed. New York: Wiley.

Graybill, F. 2000. *Theory and Application of the Linear Model*. Boston: Duxbury Press.

Meyers, R. H. 1990. *Classical and Modern Regression with Applications*, 2d ed. Boston: PWS-Kent.

Meyers, R. H., and J. S. Milton. 1998. *A First Course in the Theory of Linear Statistical Models*. New York: McGraw-Hill, Primis Custom Pub.

Montgomery, D. C., E. A. Peck, and G. G. Vining. 2006. *Introduction to Linear Regression Analysis*, 4th ed. New York: Wiley Interscience.

Ejercicios complementarios

11.95 A temperaturas que se aproximan al cero absoluto (−273°C), el helio exhibe características que desafían a numerosas leyes de la física convencional. Se realizó un experimento con helio en forma sólida a diferentes temperaturas cercanas al cero absoluto. El helio sólido se coloca en un refrigerador de dilución junto con una sustancia impura sólida y se registró la fracción (en peso) de la impureza que pasó por el helio sólido. (El fenómeno de sólidos que pasan directamente por sólidos se conoce como *efecto cuán-*

tico de túnel.) Los datos se proporcionan en la siguiente tabla.

Temperatura (x) en °C	Proporción de impureza que pasa por el helio (y)
−262.0	.315
−265.0	.202
−256.0	.204
−267.0	.620
−270.0	.715
−272.0	.935
−272.4	.957
−272.7	.906
−272.8	.985
−272.9	.987

a Ajuste una recta de mínimos cuadrados a los datos.

b Pruebe la hipótesis nula $H_0: \beta_1 = 0$ contra la hipótesis alternativa $H_a: \beta_1 < 0$, con un nivel de significancia de $\alpha = .01$.

c Encuentre un intervalo de predicción de 95% para el porcentaje de la impureza sólida que pasa por el helio sólido a -273°C. (Este valor de x está fuera de la región experimental donde el uso del modelo para predicción puede ser peligroso.)

11.96 Se realizó un estudio para determinar si existe una relación lineal entre la resistencia a la ruptura y de vigas de madera y la gravedad específica x de la madera. Diez vigas seleccionadas al azar, de iguales dimensiones de sección transversal, se sometieron a esfuerzo hasta romperlas. Las resistencias a la ruptura y la densidad de la madera se muestran en la siguiente tabla para cada una de las diez vigas.

Viga	Gravedad específica (x)	Resistencia (y)
1	.499	11.14
2	.558	12.74
3	.604	13.13
4	.441	11.51
5	.550	12.38
6	.528	12.60
7	.418	11.13
8	.480	11.70
9	.406	11.02
10	.467	11.41

a Ajuste el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$.

b Pruebe $H_0: \beta_1 = 0$ contra la hipótesis alternativa, $H_a: \beta_1 \neq 0$.

c Calcule la resistencia media para vigas con gravedad específica .590 usando un intervalo de confianza de 90%.

11.97 Una respuesta Y es una función de tres variables independientes x_1, x_2 y x_3 que están relacionadas como sigue:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$$

a Ajuste el modelo a los $n = 7$ datos de la siguiente tabla.

y	x_1	x_2	x_3
1	−3	5	−1
0	−2	0	1
0	−1	−3	1
1	0	−4	0
2	1	−3	−1
3	2	0	−1
3	3	5	1

b Prediga Y cuando $x_1 = 1, x_2 = -3, x_3 = -1$. Compare con la respuesta observada en los datos originales. ¿Por qué no son iguales los dos valores?

c ¿Los datos presentan suficiente evidencia para indicar que x_3 contribuye con información para la predicción de Y? (Pruebe la hipótesis $H_0: \beta_3 = 0$, usando $\alpha = .05$.)

d Encuentre un intervalo de confianza de 95% para el valor esperado de Y, dadas $x_1 = 1$, $x_2 = -3$ y $x_3 = -1$.

e Encuentre un intervalo de predicción de 95% para Y, dadas $x_1 = 1$, $x_2 = -3$ y $x_3 = -1$.

11.98 Si valores de las variables independientes están igualmente espaciados, ¿cuál es la ventaja de codificar nuevas variables que representen espaciamiento simétrico alrededor del origen?

11.99 Supongamos que usted desea ajustar una recta a un conjunto de n puntos, donde n es un entero par y que puede seleccionar los n valores de x del intervalo $-9 \leq x \leq 9$. ¿Cómo debería seleccionar los valores de x para minimizar $V(\hat{\beta}_1)$?

11.100 Consulte el Ejercicio 11.99. Es común emplear igual separación al seleccionar los valores de x. Suponga que $n = 10$. Encuentre la eficiencia relativa del estimador $\hat{\beta}_1$ basado en igual separación contra el mismo estimador basado en la separación del Ejercicio 11.99. Suponga que $-9 \leq x \leq 9$.

11.101 Los datos de la siguiente tabla provienen de la comparación de los porcentajes de crecimiento para bacterias tipos A y B. El crecimiento Y registrado en cinco puntos igualmente espaciados (y codificados) de tiempo se muestra en la tabla.

	Tiempo				
Tipo de bacteria	−2	−1	0	1	2
A	8.0	9.0	9.1	10.2	10.4
B	10.0	10.3	12.2	12.6	13.9

a Ajuste el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

a los $n = 10$ puntos. Sea $x_1 = 1$ si el punto se refiere a bacterias tipo B y sea $x_1 = 0$ si el punto se refiere al tipo A. Sea $x_2 =$ tiempo codificado.

b Grafique los puntos y las dos rectas de crecimiento. Observe que β_3 es la diferencia entre las pendientes de las dos rectas y representa la interacción tiempo-bacteria.

c Prediga el crecimiento del tipo A en el tiempo $x_2 = 0$ y compare la respuesta con la gráfica. Repita el proceso para el tipo B.

d ¿Los datos presentan suficiente evidencia para indicar una diferencia en los porcentajes de crecimiento para los dos tipos de bacteria?

e Encuentre un intervalo de confianza de 90% para el crecimiento esperado para el tipo B en el tiempo $x_2 = 1$.

f Encuentre un intervalo de predicción de 90% para el crecimiento Y del tipo B en el tiempo $x_2 = 1$.

11.102 El siguiente modelo fue propuesto para probar si había evidencia de discriminación salarial contra mujeres en el sistema de una universidad estatal:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_2^2 + \varepsilon,$$

donde

$$\begin{aligned} Y &= \text{salario anual (en miles de dólares)}, \\ x_1 &= \begin{cases} 1, & \text{si es mujer}, \\ 0, & \text{si es hombre}, \end{cases} \\ x_2 &= \text{cantidad de experiencia (en años)}. \end{aligned}$$

Cuando este modelo se ajustó a datos obtenidos de los registros de 200 miembros de la facultad, SSE = 783.90. El modelo reducido $Y = \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \varepsilon$ también se ajustó y produjo un valor de SSE = 795.23. ¿Los datos dan suficiente evidencia para apoyar el dicho de que el salario medio depende del género de los miembros de la facultad? Use $\alpha = .05$.

11.103 Demuestre que la ecuación de predicción de mínimos cuadrados

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

pasa por el punto $(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k, \overline{y})$.

11.104 Se realizó un experimento para determinar el efecto de presión y temperatura en el rendimiento de una sustancia química. Se emplearon dos niveles de presión (en libras por pulgada cuadrada, psi) y tres de temperatura:

Presión (psi)	Temperatura (°F)
50	100
80	200
	300

Una prueba del experimento en cada combinación de temperatura-presión dio como resultado los datos que aparecen en la siguiente tabla.

Rendimiento	Presión (psi)	Temperatura (°F)
21	50	100
23	50	200
26	50	300
22	80	100
23	80	200
28	80	300

a Ajuste el modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \varepsilon$, donde x_1 = presión y x_2 = temperatura.

b Compruebe si β_3 difiere considerablemente de cero, con $\alpha = .05$.

c Pruebe la hipótesis de que la temperatura no afecta el rendimiento, con $\alpha = .05$.

***11.105** Considere que (X, Y) tienen una distribución normal bivariante. Una prueba de $H_0: \rho = 0$ contra $H_a: \rho \neq 0$ se puede deducir como sigue.

a Sean $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$ y $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$. Demuestre que

$$\hat{\beta}_1 = r\sqrt{\frac{S_{yy}}{S_{xx}}}.$$

b Con la condición $X_i = x_i$ para $i = 1, 2, \ldots, n$, demuestre que dada $H_0: \rho = 0$

$$\frac{\hat{\beta}_1\sqrt{(n-2)S_{xx}}}{\sqrt{S_{yy}(1-r^2)}}$$

tiene una distribución t con $(n-2)$ grados de libertad.

c Con la condición $X_i = x_i$ para $i = 1, 2, \ldots, n$, concluimos que

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

tiene una distribución t con $(n-2)$ grados de libertad, dada $H_0: \rho = 0$. En consecuencia, concluimos que T tiene la misma distribución incondicionalmente.

11.106 Los costos de mano de obra y materiales son dos componentes básicos en el costo de construcción. Los cambios en los costos componentes por supuesto que llevan a cambios en los costos totales de construcción. La siguiente tabla da seguimiento a cambios en el costo de construcción y en el costo de todos los materiales de construcción durante 8 meses consecutivos.

Mes	Costo de construcción (y)	Índice de todos los materiales de construcción (x)
Enero	193.2	180.0
Febrero	193.1	181.7
Marzo	193.6	184.1
Abril	195.1	185.3
Mayo	195.6	185.7
Junio	198.1	185.9
Julio	200.9	187.7
Agosto	202.7	189.6

¿Los datos aportan suficiente evidencia para indicar una correlación diferente de cero entre los costos mensuales de construcción y los índices de todos los materiales de construcción? Calcule el nivel de significancia alcanzado.

11.107 Los datos de la siguiente tabla proporcionan las millas por galón recorridas por un automóvil de prueba cuando utiliza gasolinas de niveles variables de octanaje.

Millas por galón (y)	Octano (x)
13.0	89
13.2	93
13.0	87
13.6	90
13.3	89
13.8	95
14.1	100
14.0	98

a Calcule el valor de r.

b ¿Los datos dan suficiente evidencia para indicar que el nivel de octanaje y las millas por galón son dependientes? Obtenga el nivel de significancia alcanzado, e indique su conclusión si desea realizar una prueba en el nivel de $\alpha = .05$.

11.108 **Ejercicio Applet** Entre a la aplicación *Removing Points from Regression.* A veces, remover un punto de los que se usan para ajustar un modelo de regresión produce un modelo ajustado que es muy distinto del obtenido cuando se usan todos los datos (por ejemplo un punto recibe el nombre de punto de *alta influencia*).

a La gráfica superior proporciona un conjunto de datos y una recta de regresión ajustada útiles para predecir el peso de un estudiante según su estatura. Haga clic en cualesquiera puntos para eliminarlos y reajustar el modelo de regresión. ¿Puede hallar un punto de alta influencia?

b Arrastre a la segunda gráfica que relaciona una calificación cuantitativa SAT (Scholastic Aptitud Test) con la clasificación secundaria. ¿La pendiente de la recta de regresión ajustada le sorprende? ¿Puede hallar un punto de alta influencia? ¿Eliminar ese punto produce una recta de regresión que satisface mejor su expectativa con respecto a la relación entre calificaciones cuantitativas SAT y clasificación en el grupo?

c Arrastre al resto de los conjuntos de datos y explore qué ocurre cuando se remueven diferentes puntos.

CAPÍTULO 12

Consideraciones al diseñar experimentos

12.1 Los elementos que afectan la información en una muestra

Una medida significativa de la información que contiene una muestra para hacer una inferencia acerca de un parámetro poblacional es proporcionada por el ancho (o semiancho) del intervalo de confianza que pudiera construirse a partir de los datos muestrales. Recuerde que un intervalo de confianza de muestra grande de 95% para una media poblacional es

$$\overline{Y} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right).$$

Los anchos de muchos de los intervalos de confianza que se emplean comúnmente, como el intervalo de confianza para una media poblacional, dependen de la varianza poblacional σ^2 y el tamaño muestral n. Cuanto menor sea la variación en la población, medida por σ^2, más corto será el intervalo de confianza. Del mismo modo, el ancho del intervalo de confianza disminuye cuando n aumenta. Este interesante fenómeno nos llevaría a pensar que dos factores afectan la cantidad de información en una muestra relacionada con un parámetro: es decir, la variación de los datos y el tamaño muestral n. Encontraremos esta deducción demasiado simple pero es esencialmente verdadera.

En capítulos anteriores, cuando estuvimos interesados en comparar dos medias poblacionales o ajustar una regresión lineal simple, supusimos que muestras aleatorias independientes se tomaron de las poblaciones de interés. Si deseamos comparar dos poblaciones con base en un total de n observaciones, ¿cuántas observaciones deben tomarse de cada población?

Si hemos decidido ajustar un modelo de regresión lineal simple y deseamos maximizar la información de los datos resultantes, ¿cómo debemos escoger los valores de la variable independiente? Estas preguntas se responden en la siguiente sección.

Generalmente el diseño de experimentos es un tema muy amplio que se refiere a métodos de muestreo para reducir la variación en un experimento y, por tanto, para adquirir una cantidad especificada de información al mínimo costo. Si la meta es hacer una comparación de dos medias poblacionales, con frecuencia es suficiente el *experimento de observaciones pareadas*. Después de considerar el experimento de observaciones pareadas de la Sección 12.3, el resto del capítulo presenta algunas consideraciones importantes acerca del diseño de buenos experimentos.

12.2 Diseño de experimentos para aumentar la precisión

Como veremos, para el mismo número total de observaciones, algunos métodos de recolección de datos (*diseños*) proporcionan más información respecto a parámetros poblacionales específicos que otros. Ningún diseño individual es mejor para adquirir información respecto a todos los tipos de parámetros poblacionales. De hecho, el problema de hallar el mejor diseño para concentrar información en un parámetro poblacional específico se ha resuelto en sólo unos cuantos casos. El propósito de esta sección no es presentar una teoría general sino, más bien, presentar dos ejemplos que ilustran los principios involucrados.

Considere el problema de calcular la diferencia entre un par de medias poblacionales, $\mu_1 - \mu_2$, con base en muestras aleatorias independientes. Si el experimentador tiene recursos suficientes para muestrear un total de n observaciones, ¿cuántas observaciones debe seleccionar de las poblaciones 1 y 2, es decir, n_1 y n_2 ($n_1 + n_2 = n$), respectivamente, para maximizar la información de los datos pertinentes a $\mu_1 - \mu_2$? Si $n = 10$, ¿debe seleccionar $n_1 = n_2 = 5$ observaciones de cada población, o sería mejor una asignación de $n_1 = 4$ y $n_2 = 6$?

Si las muestras aleatorias se sacan independientemente, calculamos $\mu_1 - \mu_2$ con $\overline{Y}_1 - \overline{Y}_2$, que tiene error estándar

$$\sigma_{(\overline{Y}_1 - \overline{Y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Cuanto menor sea $\sigma_{(\overline{Y}_1 - \overline{Y}_2)}$, menor será el correspondiente error de cálculo y mayor será la cantidad de información de la muestra pertinente a $\mu_1 - \mu_2$. Si, como es frecuente suponer, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, entonces

$$\sigma_{(\overline{Y}_1 - \overline{Y}_2)} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Usted puede verificar que esta cantidad es mínima cuando $n_1 = n_2$ y, en consecuencia, que la muestra contiene un máximo de información acerca de $\mu_1 - \mu_2$ cuando las n unidades experimentales se dividen por igual entre los dos tratamientos. Un caso más general se considera en el Ejemplo 12.1.

EJEMPLO **12.1** Si han de usarse n observaciones para calcular $\mu_1 - \mu_2$, con base en muestras aleatorias independientes de las dos poblaciones de interés, encuentre n_1 y n_2 para que $V(\overline{Y}_1 - \overline{Y}_2)$ sea minimizada (suponga que $n_1 + n_2 = n$).

Solución Denote con b la fracción de las n observaciones asignadas a la muestra de la población 1; esto es, $n_1 = bn$ y $n_2 = (1 - b)n$. Entonces,

$$V(\overline{Y}_1 - \overline{Y}_2) = \frac{\sigma_1^2}{bn} + \frac{\sigma_2^2}{(1-b)n}.$$

Para hallar la fracción b que minimiza esta varianza, igualamos a cero la primera derivada con respecto a b. Este proceso da

$$-\frac{\sigma_1^2}{n}\left(\frac{1}{b^2}\right) + \frac{\sigma_2^2}{n}\left(\frac{1}{1-b}\right)^2 = 0.$$

Si despejamos b tendremos

$$b = \frac{\sigma_1}{\sigma_1 + \sigma_2} \quad \text{y} \quad 1 - b = \frac{\sigma_2}{\sigma_1 + \sigma_2}.$$

Por tanto, $V(\overline{Y}_1 - \overline{Y}_2)$ está minimizada cuando

$$n_1 = \left(\frac{\sigma_1}{\sigma_1 + \sigma_2}\right)n \quad \text{y} \quad n_2 = \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right)n,$$

es decir, cuando tamaños muestrales se asignan de manera proporcional a tamaños de las desviaciones estándar. Observe que $n_1 = n/2 = n_2$ si $\sigma_1 = \sigma_2$. ■

Como segundo ejemplo considere el problema de ajustar una recta que pasa por un conjunto de n puntos usando el método de mínimos cuadrados del Capítulo 11 (vea Figura 12.1). Además, suponga que estamos interesados principalmente en la pendiente β_1 de la recta del modelo lineal

$$Y = \beta_0 + \beta_1 x + \varepsilon.$$

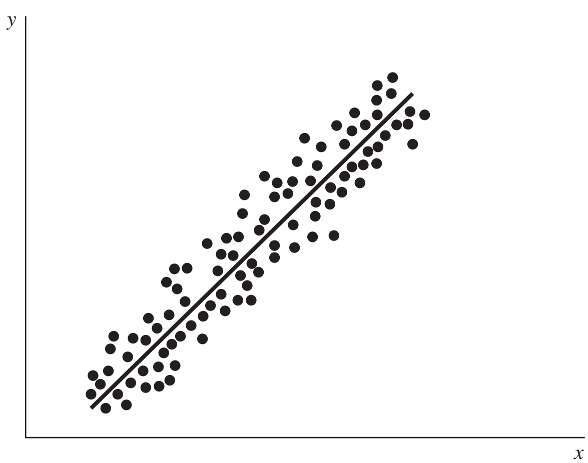


FIGURA **12.1** Ajuste de una recta por el método de mínimos cuadrados

Si tenemos la opción de seleccionar los n valores de x para los cuales y se observará, ¿cuáles valores de x maximizan la cantidad de información acerca de β_1? Tenemos una variable independiente cuantitativa x, y nuestro problema es decidir sobre los valores $x_1, x_2, \ldots, x_n$ a emplear, así como el número de observaciones a tomar en cada uno de estos valores.

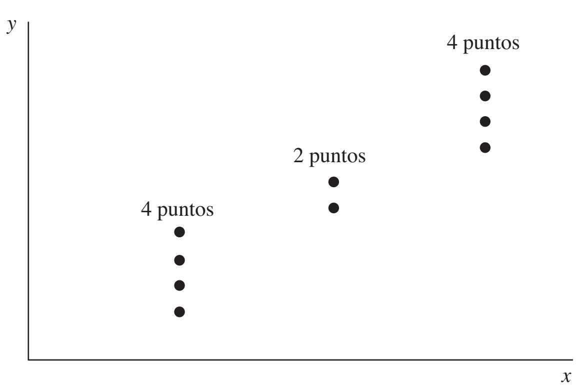
El mejor diseño para estimar la pendiente β_1 se puede determinar si consideramos la desviación estándar de $\hat{\beta}_1$:

$$\sigma_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{S_{xx}}} = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

Cuanto mayor sea S_{xx}, la suma de los cuadrados de las desviaciones de $x_1, x_2, \ldots, x_n$ alrededor de su media, menor será la desviación estándar de $\hat{\beta}_1$. Esto es, obtenemos un mejor estimador para la pendiente si los valores de x están más dispersos. En algunos casos el experimentador tiene alguna región experimental, por ejemplo $x_1 < x < x_2$, sobre la cual desea observar Y, y esta amplitud se selecciona con frecuencia antes del experimento. Entonces el valor más pequeño para $\sigma_{\hat{\beta}_1}$ se presenta cuando los n puntos están igualmente divididos, con la mitad ubicada en la frontera inferior x_1 de la región y la mitad en la frontera superior x_2. (Se omite la prueba.) Un experimentador que desee ajustar una recta, con el uso de $n = 10$ puntos en el intervalo $2 \leq x \leq 6$, seleccionaría cinco puntos en $x = 2$ y cinco en $x = 6$. Antes de concluir el análisis de este ejemplo, debe advertir que observar todos los valores de Y en sólo dos valores de x no dará información sobre la curvatura de la curva de respuesta en caso de que la suposición de linealidad en la relación de $E(Y)$ y x sea incorrecta. A menudo es más seguro seleccionar unos cuantos puntos (uno o dos) en algún lugar cercano a la parte media de la región experimental, para detectar la curvatura, si existe (véase la Figura 12.2). Un comentario adicional es oportuno. Una de las suposiciones que hemos hecho respecto al modelo de regresión lineal simple es que la varianza del término de error ε no depende del valor de la variable independiente x. Si los valores x están más dispersos, la validez de esta suposición puede ser más cuestionable.

Para resumir, hemos dado buenos diseños (asignación de unidades experimentales por población y selección de ajustes para la variable independiente x) para comparar un par de medias y ajustar una recta. Estos dos diseños simples ilustran el modo en que la información de un experimento puede aumentarse o disminuirse, dependiendo de dónde se hagan las ob-

FIGURA **12.2**
Un buen diseño para ajustar una recta ($n = 10$)



servaciones y de la asignación de tamaños muestrales. En la siguiente sección consideramos un método para controlar la cantidad de variabilidad inherente a un experimento.

Ejercicios

12.1 Supongamos que usted desea comparar las medias para dos poblaciones y que $\sigma_1^2 = 9$, $\sigma_2^2 = 25$ y $n = 90$. ¿Qué asignación de $n = 90$ para las dos muestras resultará en la máxima cantidad de información acerca de $(\mu_1 - \mu_2)$?

12.2 Consulte el Ejercicio 12.1. Supongamos que usted asigna $n_1 = n_2$ observaciones a cada muestra. ¿Qué tan grandes deben ser n_1 y n_2 para obtener la misma cantidad de información que la implicada por la solución del Ejercicio 12.1?

12.3 Suponga, como en el Ejercicio 12.1, que dos poblaciones tienen varianzas respectivas $\sigma_1^2 = 9$ y $\sigma_2^2 = 25$. Encuentre el mínimo tamaño muestral y la correspondiente asignación muestral que dará un intervalo de confianza de 95% para $\mu_1 - \mu_2$ que mida 2 unidades de longitud.

12.4 Consulte el Ejercicio 12.3. ¿Cuántas observaciones son necesarias para que un intervalo de confianza de 95% sea de 2 unidades de longitud si $n_1 = n_2$?

12.5 Suponga que deseamos estudiar el efecto del estimulante denominado digitalina en la presión sanguínea Y de ratas con dosis de $x = 2$ a $x = 5$ unidades. Se espera que la respuesta sea lineal en la región; esto es, $Y = \beta_0 + \beta_1 x + \varepsilon$. Se dispone de seis ratas para el experimento y cada rata puede recibir sólo una dosis. ¿Qué dosis de digitalina debe emplearse en el experimento y cuántas ratas deben inocularse en cada dosis para maximizar la cantidad de información del experimento con respecto a la pendiente β_1?

12.6 Consulte el Ejercicio 12.5. Considere dos métodos para seleccionar las dosis. El método 1 asigna tres ratas a la dosis $x = 2$ y tres ratas a $x = 5$. El método 2 igualmente separa las dosis entre $x = 2$ y $x = 5$ ($x = 2, 2.6, 3.2, 3.8, 4.4$ y 5.0). Suponga que σ se conoce y que la relación entre $E(Y)$ y x es verdaderamente lineal (vea el Capítulo 11). Si usamos los datos de ambos métodos para construir intervalos de confianza para la pendiente β_1, ¿cuál método dará como resultado el intervalo más largo? ¿Cuánto más largo es el intervalo? Si usamos el método 2, ¿aproximadamente cuántas observaciones serán necesarias para obtener un intervalo de la misma longitud que el obtenido por la asignación óptima del método 1?

12.7 Consulte el Ejercicio 12.5. ¿Por qué podría ser aconsejable asignar uno o dos puntos en $x = 3.5$?

12.8 El error estándar del estimador $\hat{\beta}_1$ en un modelo de regresión lineal simple se hace más pequeño cuando S_{xx} aumenta, es decir, cuando los valores x se hacen más dispersos. ¿Por qué no siempre dispersamos los valores x tanto como es posible?

12.3 El experimento de observaciones pareadas

En los Capítulos 8 y 10 consideramos métodos para comparar las medias de dos poblaciones con base en muestras independientes de cada una. En la sección anterior examinamos la forma de determinar los tamaños de las muestras de las dos poblaciones para que el error estándar del estimador $\overline{Y}_1 - \overline{Y}_2$ sea minimizado. En muchos experimentos las muestras son pareadas más que independientes. Una situación que ocurre comúnmente es aquella donde se hacen observaciones repetidas en la misma unidad de muestreo, por ejemplo obtener el peso de una misma persona antes y después de que participe en un programa de reducción de peso. En un experimento médico podríamos parear individuos que sean del mismo género y tengan pesos y edades similares. Un individuo de cada par se selecciona al azar para que reciba uno de dos medicamentos competidores para controlar la hipertensión, mientras que el otro individuo del mismo par recibe el otro medicamento.

La comparación de dos poblaciones con base en datos pareados puede ser un diseño experimental muy eficaz que puede controlar fuentes extrañas de variabilidad y resultar en la disminución del error estándar del estimador $\overline{Y}_1 - \overline{Y}_2$ para la diferencia en las medias poblacionales $\mu_1 - \mu_2$. Con (Y_{1i}, Y_{2i}), para $i = 1, 2, \ldots n$, denote una muestra aleatoria de observaciones pareadas. Suponga que

$$
\begin{aligned}
E(Y_{1i}) &= \mu_1, \qquad \mathrm{Var}(Y_{1i}) = \sigma_1^2, \qquad E(Y_{2i}) = \mu_2, \\
\mathrm{Var}(Y_{2i}) &= \sigma_2^2, \qquad \mathrm{Cov}(Y_{1i}, Y_{2i}) = \rho\sigma_1\sigma_2,
\end{aligned}
$$

donde ρ es el coeficiente de correlación común de las variables dentro de cada par (vea la Sección 5.7). Defina con $D_i = Y_{1i} - Y_{2i}$, para $i = 1, 2, \ldots n$, las diferencias entre las observaciones dentro de cada par. Debido a que los pares de observaciones se supusieron independientes y distribuidos idénticamente, los valores D_i, para $i = 1, 2, \ldots n$, son independientes y distribuidos idénticamente; usando el Teorema 5.12, vemos que

$$
\begin{aligned}
\mu_D &= E(D_i) = E(Y_{1i}) - E(Y_{2i}) = \mu_1 - \mu_2, \\
\sigma_D^2 &= \mathrm{Var}(D_i) = \mathrm{Var}(Y_{1i}) + \mathrm{Var}(Y_{2i}) - 2\mathrm{Cov}(Y_{1i}, Y_{2i}) \\
&= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2.
\end{aligned}
$$

A partir de estas consideraciones, un estimador natural para $\mu_1 - \mu_2$ es el promedio de las diferencias $\overline{D} = \overline{Y}_1 - \overline{Y}_2$, y

$$
\begin{aligned}
E(\overline{D}) &= \mu_D = \mu_1 - \mu_2, \\
\sigma_{\overline{D}}^2 &= \mathrm{Var}(\overline{D}) = \frac{\sigma_D^2}{n} = \frac{1}{n}\left[\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\right].
\end{aligned}
$$

Si la información se había obtenido de un experimento con muestras *independientes* y $n_1 = n_2 = n$,

$$
\begin{aligned}
E(\overline{Y}_1 - \overline{Y}_2) &= \mu_1 - \mu_2, \\
\sigma_{(\overline{Y}_1 - \overline{Y}_2)}^2 &= \frac{1}{n}\left[\sigma_1^2 + \sigma_2^2\right].
\end{aligned}
$$

Si es razonable creer que en los pares (Y_{1i}, Y_{2i}), para $i = 1, 2, \ldots n$, los valores de Y_{1i} y Y_{2i} tenderán a aumentar o disminuir juntos ($\rho > 0$), entonces un análisis de las expresiones anteriores para $\sigma_{\overline{D}}^2$ del experimento de pares acoplados y $\sigma_{(\overline{Y}_1 - \overline{Y}_2)}^2$ del experimento de muestras independientes demuestra que el experimento de observaciones pareadas proporciona un estimador con menor varianza que el experimento de muestras independientes. En el Ejercicio 12.11 pediremos al lector que decida cuándo es que los dos experimentos generarán estimadores con la misma varianza y cuándo el experimento de muestras independientes dará como resultado el estimador con la menor varianza.

Debido a que parear muestras hace dependientes las observaciones dentro de cada par, no podemos usar los métodos que previamente se desarrollaron para comparar poblaciones con base en muestras independientes entre sí. El análisis de un experimento de observaciones pareadas utiliza las n diferencias pareadas, D_i, para $i = 1, 2, \ldots n$. Las inferencias respecto a las diferencias en las medias $\mu_1 - \mu_2$ se construyen haciendo inferencias respecto a la media de las diferencias, μ_D. Defina

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \qquad \text{y} \qquad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$$

y utilice el procedimiento apropiado de una muestra para completar la inferencia. Si el número de pares, y por tanto el número de diferencias, es grande, por ejemplo $n > 30$, se pueden emplear los métodos inferenciales de muestra grande desarrollados en los Capítulos 8 y 10. Si el número de diferencias n es pequeño y es razonable suponer que las *diferencias están distribuidas normalmente en forma aproximada*, podemos usar métodos inferenciales con base en la distribución t. Ilustramos esto con el siguiente ejemplo.

EJEMPLO **12.2** Deseamos comparar dos métodos para determinar el porcentaje de mineral de hierro en muestras de mineral. Debido a que es probable que diferencias inherentes a las muestras de mineral contribuyan con variabilidad no deseada en las mediciones que observamos, fue creado un experimento de observaciones pareadas al dividir en dos partes cada una de las 12 muestras de mineral. La mitad de cada muestra se seleccionó al azar y se sometió al método 1; la otra mitad se sometió al método 2. Los resultados se presentan en la Tabla 12.1. ¿Los datos aportan suficiente evidencia de que el método 2 arroja un porcentaje promedio más alto que el método 1? Pruebe usando $\alpha = .05$.

Solución Hemos formado las diferencias de la Tabla 12.1 tomando la medida del método 1 y restando la correspondiente medición del método 2. Si el porcentaje medio para el método 2 es mayor, entonces $\mu_D = \mu_1 - \mu_2 < 0$. Entonces, probamos

$$H_0 : \mu_D = 0 \text{ contra } H_a : \mu_D < 0.$$

Tabla 12.1 Datos para el experimento de observaciones pareadas del Ejemplo 12.2.

Muestra de mineral	Método 1	Método 2	d_i
1	38.25	38.27	−.02
2	31.68	31.71	−.03
3	26.24	26.22	+.02
4	41.29	41.33	−.04
5	44.81	44.80	+.01
6	46.37	46.39	−.02
7	35.42	35.46	−.04
8	38.41	38.39	+.02
9	42.68	42.72	−.04
10	46.71	46.76	−.05
11	29.20	29.18	+.02
12	30.76	30.79	−.03
			$\overline{d} = -.0167$

Para estos datos,

$$s_D^2 = \frac{\sum_{i=1}^{n} d_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2}{n-1} = \frac{.0112 - \frac{1}{12}(-.20)^2}{11} = .0007.$$

Si es razonable suponer que las diferencias están distribuidas normalmente, se deduce que

$$t = \frac{\overline{d} - 0}{s_D/\sqrt{n}} = \frac{-.0167}{\sqrt{.0007}/\sqrt{12}} = -2.1865$$

es el valor observado de un estadístico que dada la hipótesis nula tiene una distribución t con $n - 1 = 11$ grados de libertad (gl). Usando la Tabla 5, Apéndice 3, con $\alpha = .05$, rechazamos H_0 si $t < -1.796$. Por tanto, deducimos que existe suficiente evidencia para permitirnos concluir que el método 2 arroja un porcentaje promedio más alto que el método 1. De nuevo, usando la Tabla 5, Apéndice 3, se deduce que $.025 <$ valor $p < .05$. La aplicación *Student's t Probabilities and Quantiles* da el valor p exacto $= P(t < -2.1865) = P(t > 2.1856) = .02564$. ■

Aunque los resultados del Ejemplo 12.2 implican que los resultados del experimento son estadísticamente significativos, podemos evaluar la significancia práctica del resultado al formar un intervalo de confianza para μ_D. Si es razonable suponer que las diferencias dentro de cada par están distribuidas normalmente en forma aproximada, un intervalo de confianza de $100(1 - \alpha)\%$ para $\mu_D = \mu_1 - \mu_2$ está dado por

$$\overline{D} \pm t_{\alpha/2}\left(\frac{S_D}{\sqrt{n}}\right),$$

donde $t_{\alpha/2}$ está basada en $n - 1$ grados de libertad (recuerde que n es el número de *pares* de observaciones).

EJEMPLO **12.3** Use los datos del Ejemplo 12.2 para formar un intervalo de confianza de 95% para la diferencia en lecturas medias de porcentaje usando los métodos 1 y 2.

Solución Del Ejemplo 12.2 observamos que

$$\overline{d} = -.0167, \qquad s_D^2 = .0007, \qquad n - 1 = 11.$$

Debido a que, con 11 grados de libertad, $t_{.025} = 2.201$, el intervalo deseado es

$$-.0167 \pm (2.201)\frac{\sqrt{.0007}}{\sqrt{12}}, \quad \text{o} \quad (-.0335, +.0001).$$ ■

Los métodos anteriores basados en la distribución t se pueden emplear de manera válida si es razonable suponer que las diferencias están distribuidas normalmente. Cuando comparamos dos medias poblacionales basadas en pequeñas muestras independientes, requerimos que las varianzas poblacionales sean iguales. La validez del análisis de observaciones pareadas no requiere la suposición de varianzas poblacionales iguales. La cantidad S_D^2 proporciona un estimador insesgado para la varianza de las diferencias, σ_D^2, cualesquiera que sean los valores de σ_1^2, σ_2^2 y ρ. La prueba t de muestras independientes también requería que ambas muestras fueran tomadas de poblaciones distribuidas normalmente. Una forma en que las diferencias dentro de pares puedan adoptar una distribución normal es que Y_{1i}, para $i = 1, 2, \ldots, n$, y Y_{2i}, para $i = 1, 2, \ldots, n$, tengan una distribución normal. No obstante, es posible que las diferencias por pares se encuentren distribuidas normalmente incluso si las Y_1 y las Y_2 no lo están. El Ejercicio 12.17 presenta un ejemplo de esta situación. Entonces, la suposición de que las diferencias estén distribuidas normalmente es menos restrictiva que la suposición de que ambas poblaciones estén distribuidas normalmente.

Hemos visto que el experimento de observaciones pareadas se puede usar para disminuir la variabilidad inherente presente en los datos. Además, en muchas situaciones, las suposiciones requeridas para emplear de manera válida un análisis de observaciones pareadas son menos restrictivas que los correspondientes métodos de muestras independientes. ¿Por qué los analistas de estadística encuentran datos de observaciones pareadas? A veces el experimento de observaciones pareadas se realizó por diseño, tomando en cuenta las consideraciones previamente discutidas. Otras veces se obtuvieron datos por medio del experimento de observaciones pareadas por comodidad. Cualquiera que sea la razón para realizar un experimento de observaciones pareadas, los datos resultantes no deben ser analizados usando un método apropiado para datos obtenidos usando muestras independientes.

Recuerde que los datos de un experimento de observaciones pareadas se analizan concentrándose en las diferencias de las observaciones dentro de cada par. Así, algunos expertos en estadística prefieren referirse al experimento de observaciones pareadas como a un experimento de diferencia pareada. En la siguiente sección presentamos alguna terminología común asociada con diseños experimentales y consideramos extensiones del experimento de las muestras independientes y el experimento de observaciones pareadas.

Ejercicios

12.9 Considere los datos analizados en los Ejemplos 12.2 y 12.3.

a Suponiendo que los dos métodos empleados para analizar las muestras trabajaron razonablemente bien, ¿por qué piensa usted que las observaciones en las dos mitades de cada muestra de mineral estarán correlacionadas positivamente?

b ¿Piensa usted que deberíamos haber tomado observaciones independientes usando los dos métodos, o deberíamos haber realizado el análisis pareado contenido en el texto? ¿Por qué?

12.10 Es frecuente que se comparen dos computadoras al ejecutar un conjunto de varios programas "de comparación" y registrar la diferencia en tiempo de la unidad de procesamiento central (CPU) necesario para completar el mismo programa. Seis programas de comparación, ejecutados en dos computadoras, produjeron la siguiente tabla de tiempos del CPU (en minutos).

	Programa de comparación					
Computadora	1	2	3	4	5	6
1	1.12	1.73	1.04	1.86	1.47	2.10
2	1.15	1.72	1.10	1.87	1.46	2.15

a ¿Los datos aportan suficiente evidencia para indicar una diferencia en los tiempos medios del CPU necesarios para que las dos computadoras terminen un trabajo? Pruebe usando $\alpha = .05$.

b Obtenga los límites para el valor p asociado.

c Encuentre un intervalo de confianza de 95% para la diferencia en tiempo medio del CPU necesario para que las dos computadoras terminen un trabajo.

12.11 Cuando Y_{1i}, para $i = 1, 2, \dots, n$, y Y_{2i}, para $i = 1,2,\dots, n$, representan muestras independientes de dos poblaciones con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, hemos determinado que $\sigma^2_{(\overline{Y}_1 - \overline{Y}_2)} = (1/n)(\sigma_1^2 + \sigma_2^2)$. Si las muestras fueron pareadas y calculamos las diferencias, D_i, para $i = 1, 2, \dots, n$, determinamos que $\sigma(2/D) = (1/n)(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$.

a ¿Cuándo es $\sigma^2_{(\overline{Y}_1 - \overline{Y}_2)}$ mayor que $\sigma\,(2/D)$?

b ¿Cuándo es $\sigma^2_{(\overline{Y}_1 - \overline{Y}_2)}$ igual que $\sigma\,(2/D)$?

c ¿Cuándo es $\sigma^2_{(\overline{Y}_1 - \overline{Y}_2)}$ menor que $\sigma\,(2/D)$?

d Con base en la exposición del texto y sus respuestas a los incisos a–c, ¿cuándo sería mejor realizar el experimento de observaciones pareadas y cuándo sería mejor llevar a cabo el experimento de muestras independientes?

12.12 Consulte el Ejercicio 12.11. Suponga que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Los valores de la tabla empleados para poner en práctica una prueba de hipótesis o construir un intervalo de confianza dependen, para muestras pequeñas, del número de grados de libertad asociados con las estimaciones para σ^2 o σ_D^2.

a Suponiendo dos muestras independientes, cada una de tamaño n, y que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, ¿cuántos grados de libertad están asociados con el estimador para la varianza común σ^2?

b Suponiendo un experimento de pares acoplados formado de n pares de observaciones, ¿cuántos grados de libertad están asociados con el estimador de σ_D^2?

c Suponga que todas las suposiciones necesarias para realizar los procedimientos t de muestras independientes están satisfechas y que deseamos hallar un intervalo de confianza de 95% para la diferencia en medias. ¿Cuáles son los valores de $t_{.025}$ usados para construir intervalos de confianza para la diferencia en medias con base en las muestras independientes y los experimentos de observaciones pareadas si $n = 5$? ¿Si $n = 10$? ¿Si $n = 30$?

d Si están satisfechas todas las suposiciones necesarias para realizar los procedimientos t de muestras independientes, identifique una posible desventaja para practicar un experimento de observaciones pareadas en lugar de tomar muestras independientes.

12.13 El Ejercicio 10.76 describe un experimento dental realizado para investigar la efectividad de un enjuague oral para inhibir el crecimiento de placa en los dientes. Los sujetos se dividieron en dos grupos: un grupo utilizó un enjuague que contenía el agente antiplaca y el grupo de control utilizó un enjuague sólo con ingredientes inactivos. Otro experimento ha sido efectuado para evaluar el crecimiento de placa para personas que han empleado el enjuague con el agente antiplaca. Para cada persona del estudio, se midió el aumento de placa 4 horas después de usar el enjuague y otra vez 8 horas después. Si desea comparar el aumento medio de placa para los dos momentos diferentes, ¿implementaría un análisis con base en un procedimiento de observaciones pareadas o de muestras independientes? ¿Por qué?

12.14 Dos procedimientos para sinterizar (calentar) cobre se han de comparar al probar cada uno de ellos en seis tipos diferentes de polvo. La medición de interés es la porosidad (porcentaje de volumen debido a huecos) de cada espécimen de prueba. Los resultados de las pruebas se muestran en la siguiente tabla.

Polvo	Procedimiento I	Procedimiento II
1	21	23
2	27	26
3	18	21
4	22	24
5	26	25
6	19	16

¿Hay suficiente evidencia para decir que el procedimiento II produce valores de porosidad más altos? Proporcione límites para el valor p. ¿Qué se concluiría en el nivel $\alpha = .05$?

12.15 Un gerente de planta, para decidir si compra una máquina de diseño A o diseño B, comprueba los tiempos para completar cierto trabajo en cada máquina. Ocho técnicos se emplearon en el experimento, con cada técnico usando ambas máquinas en un orden aleatorio. Los tiempos (en segundos) necesarios para completar el trabajo se proporcionan en la siguiente tabla.

Técnico	A	B
1	32	30
2	40	39
3	42	42
4	26	23
5	35	36
6	29	27
7	45	41
8	22	21

a Realice una prueba para ver si hay una diferencia importante entre los tiempos medios para completar el trabajo, con un nivel de significancia de 5%.

b ¿Piensa usted que parear técnicos hubiera sido mejor en este caso? Explique.

c ¿Qué suposiciones son necesarias para la prueba del inciso a?

12.16 El humus es el tipo de suelo rico y altamente orgánico que sirve como medio primario de crecimiento para la vegetación en los pantanos de la Florida. Debido a la alta concentración de material orgánico, el humus puede ser destruido con el tiempo por varias causas originadas por el hombre. Miembros de la Comisión de Caza y Pesca de Florida marcaron varios lugares en los pantanos y midieron la profundidad del humus en cada lugar, operación que repitieron 6 años después. La siguiente tabla identifica una parte de los datos (dados en pulgadas) obtenidos.

Lugar	Lectura inicial	Lectura posterior	Lugar	Lectura inicial	Lectura posterior
1	34.5	31.5	9	44.0	35.2
2	44.0	37.9	10	40.5	37.2
3	37.5	35.5	11	27.0	24.7
4	27.0	23.0	12	29.5	25.8
5	37.0	34.5	13	31.5	29.0
6	40.0	31.1	14	35.0	36.8
7	47.2	46.0	15	44.0	36.5
8	35.2	31.0			

a Realice una prueba para ver si hay suficiente evidencia para indicar una disminución en la profundidad promedio del humus durante el periodo de estudio. Proponga límites del valor p asociado. ¿Qué concluiría si deseara hacer una prueba en el nivel $\alpha = .01$? (Aunque hay libertad de tomar las diferencias necesarias en el orden que se prefiera, la respuesta dada al final del libro supone que las diferencias se formaron al tomar lecturas finales menos lecturas iniciales.)

b Indique un intervalo de confianza de 95% para la diferencia en profundidades medias del humus al final y al principio del estudio. Interprete este intervalo. [Vea la observación en seguida del inciso a.]

c Indique un intervalo de confianza de 95% para la profundidad media inicial del humus en la parte de los pantanos en que se realizó el estudio.

d Repita las instrucciones del inciso c para las lecturas finales.

e ¿Qué suposiciones son necesarias para aplicar las técnicas empleadas para contestar los incisos a y b? ¿Y los incisos c y d?

12.17 Consulte el experimento de observaciones pareadas y suponga que la i-ésima medición ($i = 1, 2$), en el j-ésimo par, donde $j = 1, 2, \ldots, n$, es

$$Y_{ij} = \mu_1 + U_j + \varepsilon_{ij},$$

donde μ_i = respuesta esperada para la población i, donde $i = 1, 2$,

U_j = una variable aleatoria que está distribuida uniformemente en el intervalo $(-1, +1)$,

ε_{ij} = error aleatorio asociado con la i-ésima medición del j-ésimo par.

Suponga que las ε_{ij} son variables aleatorias normales independientes con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$ y que U_j y ε_{ij} son independientes.

a Encuentre $E(Y_{ij})$.

b Explique que las Y_{1j} para $j = 1, 2, \ldots, n$, *no* están distribuidas normalmente. (No hay necesidad de hallar realmente la distribución de los valores Y_1.)

c Demuestre que $\text{Cov}(Y_{1j}, Y_{2j}) = 1/3$, para $j = 1, 2, \ldots, n$.

d Demuestre que $D_j = Y_{1j} - Y_{2j}$ son variables aleatorias distribuidas normalmente.

e En los incisos a–d se verificó que estas diferencias dentro de cada par pueden estar distribuidas normalmente aun cuando las mediciones individuales dentro de los pares no lo estén. ¿Puede proponer otro ejemplo que ilustre este mismo fenómeno?

12.4 Algunos diseños experimentales elementales

En los Capítulos 8 y 10 consideramos métodos para comparar las medias de dos poblaciones con base en muestras aleatorias independientes obtenidas de cada una. La Sección 12.3 se refería a una comparación de dos medias poblacionales mediante el experimento de observaciones pareadas. En esta sección presentamos consideraciones generales asociadas con diseñar experimentos. En especial consideramos extensiones de las metodologías de muestras independientes y de observaciones pareadas cuando el objetivo es comparar las medias de más de dos poblaciones.

Suponga que deseamos comparar cinco técnicas de enseñanza, A, B, C, D y E, y que usamos 125 estudiantes en el estudio. La meta es comparar las calificaciones medias en un examen estandarizado para estudiantes instruidos con cada uno de los cinco métodos. ¿Cómo se procedería? Aunque los 125 estudiantes de alguna manera sean representativos de los estudiantes a quienes están dirigidos estos métodos de enseñanza, ¿todos los estudiantes son idénticos? Obviamente la respuesta es negativa.

Es probable que haya muchachos y muchachas en el grupo y que los métodos puedan no ser igualmente eficaces para ambos géneros; también es probable que haya diferencias en las capacidades naturales de los estudiantes del grupo, lo cual resulta en que algunos estudiantes lo hagan mejor cualquiera que sea el método de enseñanza que se utilice. Diferentes estudiantes pueden provenir de familias que ponen especial atención en la educación y esto podría tener impacto en las calificaciones en el examen estandarizado. Además, puede haber otras diferencias entre los 125 estudiantes que tendrían un efecto no anticipado en las calificaciones del examen.

Con base en estas consideraciones decidimos que podría ser mejor *asignar al azar* 25 estudiantes a cada uno de cinco grupos. Cada grupo recibirá enseñanza usando una de las técnicas motivo de estudio. La división aleatoria de los estudiantes en los cinco grupos logra dos objetivos. Primero, eliminamos el posible efecto de sesgo de las características individuales de los estudiantes en las mediciones que hacemos. En segundo término proporciona una base probabilística para la selección de la muestra que permita al experto en estadística calcular probabilidades asociadas con las observaciones de la muestra y usarlas para hacer inferencias.

El experimento anterior ilustra los componentes básicos del diseño de un experimento. Las unidades experimentales del estudio son los estudiantes.

DEFINICIÓN 12.1 Las *unidades experimentales* son los objetos sobre los que se toman mediciones.

Este experimento comprende un solo *factor*, es decir, un método de enseñanza. En este experimento el factor tiene cinco *niveles*: A, B, C, D y E.

DEFINICIÓN 12.2 Los *factores* son variables controladas completamente por el experimentador. El nivel de intensidad (subcategoría distinta) de un factor es su *nivel.*

En un experimento de un solo factor como el anterior, cada nivel del factor individual representa un *tratamiento.* Así, en nuestro ejemplo de educación, hay cinco tratamientos, uno correspondiente a cada uno de los métodos de enseñanza. Como otro ejemplo, considere un experimento realizado para investigar el efecto de varias cantidades de nitrógeno y fosfato en la producción de una variedad de maíz. Una *unidad experimental* sería una superficie especificada, por ejemplo 1 acre, de maíz. Un *tratamiento* sería un número fijo de libras de nitrógeno x_1 y de fosfato x_2 aplicados a un acre determinado de maíz. Por ejemplo, un tratamiento podría ser usar $x_1 = 100$ libras de nitrógeno por acre y $x_2 = 200$ libras de fosfato. Un segundo tratamiento podría corresponder a $x_1 = 150$ y $x_2 = 100$. Observe que el experimentador podría usar diferentes cantidades (x_1, x_2) de nitrógeno y fosfato y que cada *combinación* representaría un tratamiento diferente.

DEFINICIÓN 12.3 Un *tratamiento* es una combinación específica de niveles de factor.

El experimento anterior para comparar los métodos de enseñanza A, B, C, D y E supuso dividir *al azar* los 125 estudiantes en cinco grupos, cada uno de tamaño 25. Cada grupo recibió exactamente uno de los tratamientos. Éste es un ejemplo de un diseño completamente aleatorizado.

DEFINICIÓN 12.4 Un *diseño completamente aleatorizado* para comparar k tratamientos es aquel en el que un grupo de n unidades experimentales relativamente homogéneas se dividen al azar en k subgrupos de tamaños $n_1, n_2, \ldots, n_k$ (donde $n_1 + n_2 + \cdots + n_k = n$). Todas las unidades experimentales de cada subgrupo reciben el mismo tratamiento, con cada tratamiento aplicado a exactamente un subgrupo.

Asociada a cada tratamiento está una población (a veces conceptual) consistente en todas las observaciones que habrían resultado si el tratamiento se aplicara en forma repetida. En el ejemplo de la enseñanza podríamos ver una población de todas las posibles calificaciones de examen si *todos* los estudiantes recibieran la enseñanza usando el método A. Las poblaciones conceptuales correspondientes están asociadas a cada uno de los otros métodos de enseñanza. Así, cada tratamiento tiene una población correspondiente de mediciones. Las observaciones obtenidas de un diseño completamente aleatorizado se ven de manera típica como *muestras aleatorias independientes* tomadas de las poblaciones correspondientes a cada uno de los tratamientos.

Suponga que deseamos comparar cinco marcas de aspirina, A, B, C, D y E, respecto a la cantidad media de ingrediente activo por tableta para cada una de las marcas. Decidimos seleccionar 100 tabletas al azar de la producción de cada uno de los fabricantes y usamos los resultados para realizar la comparación. En este caso, físicamente muestreamos cinco poblaciones diferentes. Aun cuando no "aplicamos" los diferentes tratamientos a un lote homogéneo de tabletas en blanco, es común referirse a este experimento como a uno que comprende un solo factor (el fabricante) y cinco tratamientos (correspondientes a los diferentes fabricantes). Entonces, en este ejemplo, para cada población, identificamos un tratamiento correspondiente. Ya sea que hayamos puesto en práctica un diseño completamente aleatorizado o que hayamos tomado muestras independientes de cada una de varias poblaciones existentes, se establece una correspondencia biunívoca entre las poblaciones y los tratamientos. Estas dos situaciones, en las que se toman muestras independientes de cada una de las k poblaciones, son ejemplos de un diseño biunívoco.

DEFINICIÓN **12.5**

Un *diseño biunívoco* para comparar k poblaciones es un arreglo en el que se obtienen muestras aleatorias de cada una de las poblaciones de interés.

En esta forma, un diseño biunívoco, ya sea que corresponda a datos obtenidos con el uso de un diseño completamente aleatorizado o al tomar muestras independientes de cada una de varias poblaciones existentes, es la extensión de los experimentos muestrales independientes que consideramos en los Capítulos 8 y 10. Los métodos para analizar información obtenida de un diseño biunívoco se presentan en las Secciones 13.3 a la 13.7.

En la Sección 12.3 vimos que es frecuente que un diseño de observaciones pareadas sea un método superior para comparar las medias de dos poblaciones o tratamientos. Cuando estamos interesados en comparar las efectividades de dos medicamentos para controlar la hipertensión, sugerimos formar observaciones pareadas de individuos que fueran del mismo sexo y de edades y pesos similares. Un miembro seleccionado al azar de cada par recibió el tratamiento 1, en tanto que otro recibió el tratamiento 2. La meta era controlar fuentes externas de variabilidad y así obtener un análisis más preciso. Suponga que deseamos comparar tres medicamentos diferentes y no sólo dos. ¿Cómo procederíamos? En lugar de formar varios pares de individuos de características iguales, podríamos formar varios grupos, cada uno de ellos con tres miembros igualados en cuanto a sexo, peso y edad. Dentro de cada grupo de tres, seleccionaríamos al azar un individuo para recibir el tratamiento 1 y otro para recibir el tratamiento 2 y luego administraríamos el tratamiento 3 al miembro restante de cada grupo. El objetivo de este diseño es idéntico al del diseño de observaciones pareadas, es decir, eliminar fuentes no deseadas de variabilidad que podrían entrar de manera subrepticia en las observaciones de nuestro experimento. Esta extensión del diseño de observaciones pareadas recibe el nombre de diseño aleatorizado en bloque.

DEFINICIÓN 12.6

Un *diseño aleatorizado en bloque* contiene b bloques y k tratamientos consistentes de b bloques de k unidades experimentales cada uno. Los tratamientos se asignan de manera aleatoria a las unidades de cada bloque, con cada tratamiento apareciendo exactamente una vez en cada bloque.

La diferencia entre un diseño de bloque aleatorizado y el diseño completamente aleatorizado se puede demostrar al considerar un experimento diseñado para comparar la reacción de una persona a un conjunto de cuatro estímulos (tratamientos), en un experimento psicológico de estímulo–respuesta. Denotaremos los tratamientos como T_1, T_2, T_3 y T_4.

Suponga que ocho personas se asignan al azar a cada uno de cuatro tratamientos. La asignación aleatoria de personas a tratamientos (o viceversa) distribuye al azar errores debidos a la variabilidad de persona a persona en respuesta a los cuatro tratamientos y proporciona cuatro muestras que, para todos los fines prácticos, son aleatorias e independientes. Éste es un diseño experimental completamente aleatorizado.

El error experimental asociado con un diseño completamente aleatorizado tiene varios componentes. Algunos de éstos se deben a las diferencias entre personas, a que las mediciones repetidas de una persona no sean idénticas (debido a variaciones en las condiciones físicas y psicológicas), a que el experimentador no administre un estímulo determinado con exactamente la misma intensidad en mediciones repetidas y a errores de medición. La reducción de cualquiera de estas causas de error aumentará la información del experimento.

La variación de persona a persona del experimento anterior se puede eliminar con el uso de personas como bloques. Cada persona recibiría cada uno de los cuatro tratamientos asignados en una secuencia aleatoria. El resultante diseño de bloque aleatorizado aparecería como en la Figura 12.3. Ahora sólo ocho personas son necesarias para obtener ocho mediciones de respuesta por tratamiento. Observe que cada tratamiento ocurre exactamente una vez en cada bloque.

La palabra *aleatorizado* en el nombre del diseño implica que los tratamientos son asignados al azar dentro de un bloque. Para nuestro experimento, la posición del bloque se refiere a la posición en la secuencia de estímulos asignada a una persona dada en el tiempo. El propósito de hacer aleatorio esto (es decir, la posición del bloque) es eliminar el sesgo causado por fatiga o aprendizaje.

Los bloques pueden representar tiempo, ubicación o material experimental. Si se han de comparar tres tratamientos y se sospecha de la presencia de una tendencia en la respuesta media en el tiempo, una parte importante de la variación tiempo-tendencia puede ser eliminada con bloques. Los tres tratamientos se aplicarían al azar a unidades experimentales en un pequeño bloque de tiempo. Este procedimiento se repetiría en bloques sucesivos de tiempo

FIGURA 12.3 Diseño de bloque aleatorizado

Personas

1	2	3	4	• • •	8
T_2	T_4	T_1	T_1	• • •	T_2
T_1	T_2	T_3	T_4		T_3
T_4	T_1	T_2	T_3		T_4
T_3	T_3	T_4	T_2	• • •	T_1

hasta que se recolecte la cantidad de información necesaria. Una comparación de la venta de productos competitivos en supermercados debería hacerse dentro de supermercados, usándolos como bloques y eliminando la variabilidad de tienda a tienda. Es frecuente que experimentos con animales en agricultura y medicina utilicen camadas como bloques, aplicando todos los tratamientos, uno a cada uno, a los animales dentro de la camada. Debido a características hereditarias, los animales de una camada son más homogéneos que los de varias camadas. Este tipo de bloque elimina la variación de una camada a otra. El análisis de información generada por un diseño de bloque aleatorizado se analiza en las Secciones 13.8 a 13.10.

El diseño de bloque aleatorizado es sólo uno de los muchos tipos de diseños de bloque. El diseño de bloques en dos direcciones se puede lograr con el uso de un diseño de cuadro latino. Supongamos que los individuos del ejemplo anterior se fatigaron cuando se les aplicaron los estímulos, de modo que el último estímulo siempre produjo una respuesta más baja que el primero. Si esta tendencia (y la consecuente falta de homogeneidad de las unidades experimentales dentro de un bloque) fuera verdadera para todos los individuos, un diseño de cuadro latino sería apropiado. El diseño se construiría como se indica en la Figura 12.4. Cada estímulo se aplica una vez a cada individuo y ocurre exactamente una vez en cada posición del orden de presentación. Los cuatro estímulos se presentan en cada fila (renglón) y en cada columna de la configuración de 4×4. El diseño resultante es un cuadro latino de 4×4. Un diseño de cuadro latino para tres tratamientos requiere una configuración de 3×3; en general, p tratamientos requieren un conjunto $p \times p$ de unidades experimentales. Si se desean más observaciones por tratamiento, el experimentador debe usar varias configuraciones de cuadro latino en un experimento. En el ejemplo anterior sería necesario correr dos cuadros latinos para obtener ocho observaciones por tratamiento. El experimento contendría entonces el mismo número de observaciones por tratamiento que el diseño aleatorizado de bloques (Figura 12.3).

Una comparación de medias para cualquier par de estímulos eliminaría el efecto de la variación de un individuo a otro, pero también eliminaría el efecto de la tendencia de fatiga dentro de cada estímulo porque cada tratamiento se aplicaría en cada posición de la secuencia de administrar estímulos-tiempo. Por tanto, el efecto de la tendencia se cancelaría al comparar las medias. Una exposición más amplia de diseños de bloques y sus análisis está contenida en los textos citados en la bibliografía que aparece al final del capítulo.

El objetivo de esta sección ha sido presentar algunas de las consideraciones básicas al diseñar experimentos. Hemos examinado el papel de la aleatorización en todos los experimentos bien diseñados y nos hemos concentrado en las extensiones de las muestras independientes y

FIGURA **12.4**
Diseño de cuadro latino

Orden de presentación de estímulos (filas)		1	2	3	4
	1	T_1	T_2	T_3	T_4
	2	T_2	T_3	T_4	T_1
	3	T_3	T_4	T_1	T_2
	4	T_4	T_1	T_2	T_3

experimentos de observaciones pareadas a situaciones en las que deseamos comparar más de dos tratamientos. En particular, señalamos la existencia de diseños de bloques, cómo funcionan y cómo pueden producir aumentos importantes en la cantidad de información obtenida de un experimento al reducir variaciones perjudiciales.

Ejercicios

12.18 Dos medicamentos, A y B, se han de aplicar cada uno a cinco ratas. Suponga que las ratas están numeradas del 1 al 10. Use la tabla de números aleatorios para asignar las ratas aleatoriamente a los dos tratamientos.

12.19 Consulte el Ejercicio 12.18. Suponga que el experimento comprendía tres medicamentos, A, B y C, con 5 ratas asignadas a cada uno. Use la tabla de números aleatorios para asignar aleatoriamente las 15 ratas a los tres tratamientos.

12.20 Un ingeniero químico tiene dos catalizadores y tres temperaturas que desea usar en una serie de experimentos.

a ¿Cuántos *tratamientos* (combinaciones de factor-nivel) hay en este experimento? Describa cuidadosamente uno de estos tratamientos.

b Cada experimento hace uso de una combinación de catalizador–temperatura. Muestre la forma en que usaría una tabla de números aleatorios para hacer aleatorio el orden de los experimentos.

12.21 Dé dos razones para utilizar la *aleatorización* en un experimento.

12.22 ¿Qué es un *factor*?

12.23 ¿Qué es un *tratamiento*?

12.24 ¿Una variable podría ser un factor en un experimento y variable ruidosa (fuente de variación extraña) en otro?

12.25 Si usted fuera a diseñar un experimento, ¿qué parte del procedimiento de diseño aumentaría la precisión del experimento? ¿Qué parte del procedimiento de diseño disminuiría el impacto de fuentes extrañas de variabilidad?

12.26 Se realizará un experimento para comparar el efecto de la digitalina en la contracción del músculo cardiaco de ratas. El experimento se realiza al remover el corazón de una rata viva, cortar el corazón en capas delgadas y tratar las capas con dosis de digitalina; a continuación se mide la contracción muscular. Si se van a emplear cuatro dosis, A, B, C y D, ¿qué ventaja se obtiene al aplicar A, B, C y D a una rebanada de tejido del corazón de cada rata? ¿Qué principio de diseño está ilustrado por este ejemplo?

12.27 Complete la asignación de tratamientos para el siguiente diseño de cuadro latino de 3×3.

	A	
C		

12.5 Resumen

El objetivo de este capítulo ha sido identificar los factores que afectan la cantidad de información en un experimento y usar este conocimiento para diseñar mejores experimentos. El diseño de experimentos es un tema muy amplio y ciertamente difícil de condensar en un solo capítulo de un texto de introducción. No obstante, la filosofía que sirve de base al diseño, algunos métodos para variar información en un experimento y algunas estrategias deseables para diseño se explican con facilidad.

Hemos visto que la cantidad de información relacionada con un parámetro de interés depende de la selección de combinaciones factor-nivel (tratamientos) a incluir en el experimento, así como de la asignación del número total de unidades experimentales a los tratamientos. La aleatorización es un componente importante de cualquier experimento de diseño, su uso ayuda a eliminar sesgos en resultados experimentales y proporciona la base teórica para calcular las probabilidades que son clave para el proceso de hacer inferencias. El procedimiento de bloque, es decir, comparar tratamientos dentro de bloques de material experimental relativamente homogéneos, se puede usar para eliminar variación de un bloque a otro cuando se comparan tratamientos. De esta manera, sirve como filtro para reducir el efecto de fuentes no deseadas de variabilidad.

El análisis de algunos diseños experimentales elementales se aborda en el Capítulo 13. Un tratamiento más extenso del diseño y análisis de experimentos es un curso por sí mismo. Si usted está interesado en explorar este tema, consulte los textos citados en la bibliografía recomendada.

Bibliografía y lecturas adicionales

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 2005. *Statistics for Experimenters*, 2d ed. New York: Wiley Interscience.

Cochran, W. G., and G. Cox. 1992. *Experimental Designs*, 2d ed. New York: Wiley.

Graybill, F. 2000. *Theory and Application of the Linear Model*. Belmont Calif.: Duxbury.

Hicks, C. R., and K. V. Turner. 1999. *Fundamental Concepts in the Design of Experiments*, 5th ed. New York: Oxford University Press.

Hocking, R. R. 2003. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 5th ed. New York: Wiley Interscience.

Montgomery, D. C. 2006. *Design and Analysis of Experiments*, 6th ed. New York: Wiley.

Scheaffer, R. L., W. Mendenhall, and L. Ott. 2006. *Elementary Survey Sampling*, 6th ed. Belmont Calif.: Duxbury.

Scheffé, H. 2005. *The Analysis of Variance*. New York: Wiley Interscience.

Ejercicios complementarios

12.28 ¿Cómo se puede medir la información en una muestra relacionada con un parámetro poblacional específico?

12.29 ¿Qué es una *muestra aleatoria*?

12.30 ¿Qué factores afectan la cantidad de información en un experimento? ¿Qué procedimientos de diseño controlan estos factores?

12.31 Consulte el experimento de observaciones pareadas de la Sección 12.3 y suponga que la medición que recibe el tratamiento i, donde $i = 1, 2$, en el j-ésimo par, donde $j = 1, 2, \ldots, n$, es

$$Y_{ij} = \mu_i + P_j + \varepsilon_{ij},$$

donde μ_i = respuesta esperada para el tratamiento i, para $i = 1, 2$.

P_j = contribución de efecto aleatorio aditivo (positivo o negativo) por el j-ésimo par de unidades experimentales, para $j = 1, 2, \ldots, n$.

ε_{ij} = error aleatorio asociado con la unidad experimental del j-ésimo par que recibe el tratamiento i.

Suponga que las ε_{ij} son variables aleatorias normales independientes con $E(\varepsilon_{ij}) = 0$, $V(\varepsilon_{ij}) = \sigma^2$; y suponga que las P_j son variables aleatorias normales independientes con $E(P_j) = 0$, $V(P_j) = \sigma_p^2$. También, suponga que las P_j y las ε_{ij} son independientes.

a Encuentre $E(Y_{ij})$.

b Encuentre $E(\overline{Y}_i)$ y $V(\overline{Y}_i)$, donde $\overline{Y}_i$ es la media de las n observaciones que reciben el tratamiento i, donde $i = 1, 2$.

c Sea $\overline{D} = \overline{Y}_1 - \overline{Y}_2$. Encuentre $E(\overline{D})$, $V(\overline{D})$ y la distribución de probabilidad para $\overline{D}$.

12.32 Consulte el Ejercicio 12.31. Demuestre que

$$\frac{\overline{D}\sqrt{n}}{S_D}$$

tiene una distribución t, dada $H_0 : (\mu_1 - \mu_2) = 0$.

***12.33** Consulte el Ejercicio 12.31. Suponga que un diseño completamente aleatorizado se utiliza para la comparación de las dos medias de tratamiento. Entonces, una respuesta podría ser modelada por la expresión

$$Y_{ij} = \mu_i + P_{ij} + \varepsilon_{ij},$$

pero el "efecto par" P_{ij} (que todavía afecta una unidad experimental) será seleccionado aleatoriamente y es probable que difiera de una a otra de las $2n$ observaciones que hay. Además, en contraste con el experimento de observaciones pareadas, los efectos de par no se cancelarán cuando se calcule $(\overline{Y}_1 - \overline{Y}_2)$. Compare $V(\overline{Y}_1 - \overline{Y}_2) = V(\overline{D})$ para este diseño con el diseño de observaciones pareadas del Ejercicio 12.31. ¿Por qué la varianza para el diseño completamente aleatorizado suele ser más grande?[1]

12.34 A las personas que envían trabajos de computación a un centro de cómputo por lo general se les pide estimen la cantidad de tiempo de máquina necesario para completar el trabajo. Este tiempo se mide en unidades CPU, la cantidad de tiempo que un trabajo ocupará una parte de la memoria del CPU (unidad de procesamiento central) de la computadora. Un centro de cómputo decidió efectuar una comparación de los tiempos de CPU estimado contra real para un cliente en particular. Los tiempos correspondientes estuvieron disponibles para 11 trabajos. Los datos muestrales se dan en la siguiente tabla.

1. Los ejercicios precedidos de un asterisco son opcionales.

Tiempo de CPU (minutos)	Número de trabajo 1	2	3	4	5	6	7	8	9	10	11
Estimado	.50	1.40	.95	.45	.75	1.20	1.60	2.6	1.30	.85	.60
Real	.46	1.52	.99	.53	.71	1.31	1.49	2.9	1.41	.83	.74

a ¿Por qué se esperaría que las observaciones dentro de cada uno de estos pares de datos estén correlacionadas?

b ¿Los datos dan suficiente evidencia para indicar que, *en promedio*, el cliente tiende a subestimar el tiempo de CPU necesario para calcular trabajos? Pruebe usando $\alpha = .10$.

c Encuentre el nivel de significancia observado para la prueba e interprete este valor.

d Obtenga un intervalo de confianza de 90% para la diferencia en tiempo medio estimado de CPU contra tiempo medio real de CPU.

12.35 La temperatura de la Tierra afecta la germinación de semillas, la supervivencia de cosechas en mal tiempo y muchos otros aspectos de producción agrícola. La temperatura en varios lugares se puede medir usando sensores colocados en tierra o mediante dispositivos sensibles a rayos infrarrojos instalados en aviones o satélites espaciales. La captación de información de los sensores colocados en tierra es tediosa y requiere de numerosas réplicas para obtener estimaciones precisas de la temperatura del suelo. Por otra parte, los sensores instalados en aviones o en satélites parecen introducir un sesgo en las lecturas de temperatura. Para estimar la cantidad de sesgo, ambos métodos se emplearon para medir la temperatura del suelo en cinco lugares. Las lecturas, medidas en grados Celsius, se dan en la siguiente tabla.

Lugar	Temperatura (ºC) Suelo	Aire
1	46.9	47.3
2	45.4	48.1
3	36.3	37.9
4	31.0	32.7
5	24.7	26.2

a ¿La información presenta suficiente evidencia para decir que hay diferencia en el promedio de lecturas de la temperatura en tierra usando sensores instalados en tierra y en naves?

b Construya un intervalo de confianza de 95% para la diferencia en lecturas medias de la temperatura del suelo usando sensores en tierra e instalados en las naves.

c Deseamos estimar la diferencia entre lecturas medias de la temperatura para sensores instalados en tierra y en naves a no más de .2°C en el nivel de confianza de 95%. ¿Aproximadamente cuántas observaciones pareadas (mediciones en diferentes lugares) se requieren?

12.36 Se realizó un experimento para comparar el tiempo medio de reacción a dos tipos de señalamientos de tránsito: prohibido (no vuelta a la izquierda) y permitido (sólo vuelta a la izquierda). Diez personas estuvieron incluidas en el experimento. A cada persona se le presentaron 40 señalamientos de tránsito, 20 de prohibido y 20 de permitido, en orden aleatorio. El tiempo medio de reacción y el número de acciones correctas se registraron para cada persona. Los tiempos medios de reacción a los señalamientos de tránsito, 20 de prohibido y 20 de permitido para cada una de las personas, se reproducen en la tabla siguiente.

Persona	Tiempos medios de reacción (ms) para 20 señalamientos de tránsito	
	Prohibido	Permitido
1	824	702
2	866	725
3	841	744
4	770	663
5	829	792
6	764	708
7	857	747
8	831	685
9	846	742
10	759	610

a Explique por qué éste es un experimento de observaciones pareadas y dé las razones por las cuales el pareamiento es útil para aumentar información sobre la diferencia entre los tiempos medios de reacción a señalamientos de tránsito de prohibido y de permitido.

b ¿Los datos presentan suficiente evidencia para indicar una diferencia en los tiempos medios de reacción ante señalamientos de tránsito de prohibido y de permitido? Pruebe usando $\alpha = .05$.

c Encuentre e interprete el valor p aproximado para la prueba del inciso b.

d Encuentre un intervalo de confianza de 95% para la diferencia en los tiempos medios de reacción para señalamientos de tránsito de prohibido y de permitido.

***12.37** Supongamos que usted desea ajustar el modelo

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \boldsymbol{\varepsilon}$$

a un conjunto de n datos. Si los n datos han de asignarse a los puntos de diseño $x = -1$, 0 y 1, ¿qué fracción debe asignarse a cada valor de x para minimizar $V(\hat{\beta}_2)$? (Suponga que n es grande y que k_1, k_2 y k_3, $k_1 + k_2 + k_3 = 1$, son las fracciones del número total de observaciones a ser asignadas a $x = -1$, 0 y 1, respectivamente.)

CAPÍTULO 13

El análisis de varianza

13.1 Introducción

La mayoría de los experimentos comprenden un estudio del efecto de una o más variables independientes sobre una respuesta. Las variables independientes que pueden ser controladas en un experimento reciben el nombre de *factores* y el nivel de intensidad de un factor se denomina *nivel* del factor.

El análisis de los datos generados por un experimento multivariable requiere de la identificación de las variables independientes del experimento. Éstas no serán sólo factores (variables independientes controladas) sino que también podrían ser direcciones de bloqueo. Si estudiamos las mediciones de desgaste para tres tipos de llantas, A, B y C, en cada uno de cuatro automóviles, "tipos de llantas" es un factor que representa una sola *variable cualitativa* (no hay un valor cuantitativo o numérico asociado con la variable "tipo de llanta") con tres niveles. Los automóviles son bloques y representan una sola variable cualitativa con cuatro niveles. La respuesta para un diseño de cuadro latino depende de los factores que representan tratamientos, pero también son afectadas por dos variables cualitativas independientes de bloque, "filas" y "columnas."

Los métodos para diseñar experimentos para aumentar la precisión y controlar fuentes extrañas de variación se estudiaron en el Capítulo 12. En particular, se mostró que el diseño de un factor y el diseño de bloques aleatorizado eran generalizaciones de diseños simples para las muestras independientes y comparaciones pareadas de medias que se estudiaron en los Capítulos 8, 10 y 12. Los tratamientos corresponden a combinaciones de niveles de factor e identifican las diferentes poblaciones de interés para el experimentador. Este capítulo presenta una introducción al análisis de varianza y da métodos para el análisis del diseño de un factor (incluyendo el diseño completamente aleatorizado) y diseños de bloque aleatorizados. Los métodos análogos de análisis para el diseño de cuadro latino no se presentan en este capítulo, pero pueden hallarse en los textos citados en la bibliografía al final del capítulo.

13.2 Procedimiento del análisis de varianza

El método de análisis para experimentos que comprenden varias variables independientes puede explicarse si se desarrolla de manera intuitiva el procedimiento o bien, en forma más rigurosa, por el método de modelos lineales desarrollado en el Capítulo 11. Empezamos por presentar una exposición intuitiva de un procedimiento conocido como *análisis de varianza* (ANOVA). Un resumen del método del modelo lineal se presenta en la Sección 13.13.

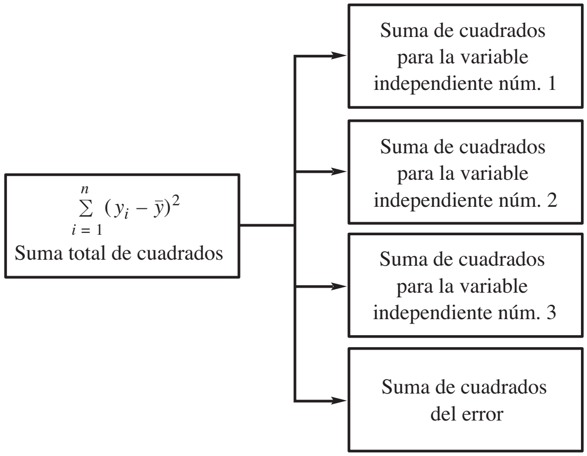
Como su nombre lo indica, el procedimiento ANOVA trata de analizar la variación en un conjunto de respuestas y asignar partes de esta variación a cada variable en un conjunto de variables independientes. Debido a que el experimentador raras veces incluye, si lo hace, todas las variables que afectan la respuesta en un experimento, la variación aleatoria en las respuestas se observa incluso si todas las variables independientes consideradas por el experimentador se mantienen constantes. El objetivo del ANOVA es identificar variables independientes importantes y determinar la forma en que afectan la respuesta.

La lógica del ANOVA puede comprenderse mejor con una explicación abstracta. El análisis real, es decir, cómo hacerlo, se ilustra con un ejemplo.

Al igual que en el Capítulo 11, la variabilidad de un conjunto de n mediciones es cuantificada por la suma de cuadrados de las desviaciones $\sum_{i=1}^{n} (y_i - \overline{y})^2$. El procedimiento ANOVA divide en partes esta suma de cuadrados de las desviaciones, llamada *suma total de cuadrados*, cada una de las cuales se atribuye a una de las variables independientes del experimento, más un residuo que está asociado con error aleatorio. La Figura 13.1 ilustra esa división para tres variables independientes. Si se escribiera un modelo lineal multivariable para la respuesta, como se sugiere en el Capítulo 11, la parte de la suma total de los cuadrados del error se denomina suma de cuadrados asignada al error (SSE).

Para los casos que consideramos y dada la hipótesis de que las variables independientes no están relacionadas con la respuesta, cada una de las partes de la suma total de cuadrados, dividida entre una constante apropiada, da como resultado un estimador independiente e insesgado de σ^2, la varianza del error experimental. Cuando una variable está altamente relacionada con la respuesta, su parte de la suma total de cuadrados (llamada *suma de cuadrados* de

FIGURA **13.1**
División de la suma total de cuadrados de las desviaciones



la variable) aumentará. Esta condición puede detectarse si se compara la suma de cuadrados para esa variable con la suma de cuadrados del error, SSE. La prueba estará basada en un estadístico que posee una distribución F y especifica que la hipótesis de que no hay efecto para la variable independiente debe ser rechazada si el valor de F es grande.

El mecanismo del ANOVA puede ilustrarse mejor si se considera un ejemplo conocido. Suponga que deseamos usar información en muestras independientes de tamaños $n_1 = n_2$ para comparar las medias de dos poblaciones distribuidas normalmente con medias μ_1 y μ_2 y varianzas iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Este experimento, ya antes analizado con el uso de la prueba t de muestras independientes, se abordará ahora desde otro punto de vista. La variación total de las mediciones de respuesta de las dos muestras es cuantificada por (recuerde que $n_1 = n_2$)

$$\text{SS total} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(Y_{ij} - \overline{Y})^2,$$

donde Y_{ij} denota la j-ésima observación de la i-ésima muestra y $\overline{Y}$ es la media de todas las $n = 2n_1$ observaciones. Esta cantidad puede dividirse en dos partes de la siguiente manera:

$$\begin{aligned}\text{SS total} &= \sum_{i=1}^{2}\sum_{j=1}^{n_1}(Y_{ij} - \overline{Y})^2 \\ &= \underbrace{n_1\sum_{i=1}^{2}(\overline{Y}_i - \overline{Y})^2}_{\text{SST}} + \underbrace{\sum_{i=1}^{2}\sum_{j=1}^{n_1}(Y_{ij} - \overline{Y}_i)^2}_{\text{SSE}}\end{aligned}$$

(prueba diferida a la Sección 13.6), donde $\overline{Y}_i$ es el promedio de las observaciones de la i-ésima muestra, para $i = 1, 2$. Examinemos la cantidad SSE más de cerca. Recuerde que hemos supuesto que las varianzas poblacionales subyacentes son iguales y que $n_1 = n_2$.

$$\begin{aligned}\text{SSE} &= \sum_{i=1}^{2}\sum_{j=1}^{n_1}(Y_{ij} - \overline{Y}_i)^2 = \sum_{i=1}^{2}(n_1 - 1)S_i^2 \\ &= (n_1 - 1)S_1^2 + (n_1 - 1)S_2^2,\end{aligned}$$

donde

$$S_i^2 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(Y_{ij} - \overline{Y}_i)^2.$$

Recuerde que, en el caso $n_1 = n_2$, el estimador "agrupado" para la varianza común σ^2 está dado por

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_1 - 1)S_2^2}{n_1 + n_1 - 2} = \frac{\text{SSE}}{2n_1 - 2}.$$

Hemos dividido la suma total de cuadrados de las desviaciones en dos partes. Una parte, SSE, puede dividirse entre $2n_1 - 2$ para obtener el estimador agrupado de σ^2. Debido a que hay sólo dos tratamientos (o poblaciones) y $n_1 = n_2$, la otra parte,

$$\text{SST} = n_1\sum_{i=1}^{2}(\overline{Y}_i - \overline{Y})^2 = \frac{n_1}{2}(\overline{Y}_1 - \overline{Y}_2)^2,$$

la *suma de cuadrados de los tratamientos* (SST), será grande si $|\overline{Y}_1 - \overline{Y}_2|$ es grande. En consecuencia, cuanto más grande sea SST, mayor será el peso de la evidencia para indicar una diferencia entre μ_1 y μ_2. ¿Cuándo será la SST suficientemente grande para indicar una diferencia importante entre μ_1 y μ_2?

Debido a que hemos supuesto que Y_{ij} está distribuida normalmente con $E(Y_{ij}) = \mu_i$, para $i = 1, 2$ y $V(Y_{ij}) = \sigma^2$ y como $\text{SSE}/(2n_1 - 2)$ es idéntica al estimador agrupado de σ^2 empleado en los Capítulos 8 y 10 se deduce que

$$E\left(\frac{\text{SSE}}{2n_1 - 2}\right) = \sigma^2$$

y que

$$\frac{\text{SSE}}{\sigma^2} = \sum_{j=1}^{n_1}\frac{(Y_{1j} - \overline{Y}_1)^2}{\sigma^2} + \sum_{j=1}^{n_1}\frac{(Y_{2j} - \overline{Y}_2)^2}{\sigma^2}$$

tiene una distribución χ^2 con $2n_1 - 2$ grados de libertad (gl) (vea la Sección 8.8).

En la Sección 13.6 obtendremos un resultado que implica que

$$E(\text{SST}) = \sigma^2 + \frac{n_1}{2}(\mu_1 - \mu_2)^2.$$

Observe que la SST estima σ^2 si $\mu_1 = \mu_2$ y una cantidad mayor que σ^2 si $\mu_1 \neq \mu_2$. Dada la hipótesis de que $\mu_1 = \mu_2$, se deduce que

$$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{2\sigma^2/n_1}}$$

tiene una distribución normal estándar; por tanto,

$$Z^2 = \left(\frac{n_1}{2}\right)\left[\frac{(\overline{Y}_1 - \overline{Y}_2)^2}{\sigma^2}\right] = \frac{\text{SST}}{\sigma^2}$$

tiene una distribución χ^2 con 1 grado de libertad.

Observe que la SST es una función sólo de las medias muéstrales $\overline{Y}_1$ y $\overline{Y}_2$ en tanto que la SSE es una función sólo de las varianzas muestrales S_1^2 y S_2^2. El Teorema 7.3 implica que, para $i = 1, 2$, las medias muestrales $\overline{Y}_i$ y las varianzas muestrales S_i^2 son independientes. Como se supone que las muestras son independientes, se deduce que la SST y la SSE son variables aleatorias independientes. En consecuencia, de la Definición 7.3, de acuerdo con la hipótesis de que $\mu_1 = \mu_2$,

$$\frac{\dfrac{\text{SST}}{\sigma^2}\Big/1}{\dfrac{\text{SSE}}{\sigma^2}\Big/(2n_1-2)} = \frac{\text{SST}/1}{\text{SSE}/(2n_1-2)}$$

tiene una distribución F con $\nu_1 = 1$ grado de libertad en el numerador y $\nu_2 = (2n_1 - 2)$ grados de libertad en el denominador.

Las sumas de cuadrados divididas entre sus respectivos grados de libertad reciben el nombre de *cuadrados medios*. En este caso, el cuadrado medio del error y el cuadrado medio de los tratamientos están dados por

$$\text{MSE} = \frac{\text{SSE}}{2n_1 - 2} \quad \text{y} \quad \text{MST} = \frac{\text{SST}}{1}.$$

de acuerdo con $H_0 : \mu_1 = \mu_2$, MST y MSE representan estimaciones de σ^2. No obstante, cuando H_0 es falsa y $\mu_1 \neq \mu_2$, MST constituye una estimación de una cantidad mayor que σ^2 y tiende a ser más grande que MSE. Para probar $H_0 : \mu_1 = \mu_2$ contra $H_a : \mu_1 \neq \mu_2$, usamos

$$F = \frac{\text{MST}}{\text{MSE}}$$

como el estadístico de prueba.

El desacuerdo con la hipótesis nula está indicado por un valor grande de F; en consecuencia, la región de rechazo para una prueba con nivel de significancia α es

$$F > F_\alpha.$$

Entonces, la prueba ANOVA resulta en una prueba F de una cola. Los grados de libertad para F son los asociados con MST y MSE. En este ejemplo, como ya antes indicamos, F está basada en $\nu_1 = 1$ y $\nu_2 = 2n_1 - 2$ grados de libertad en numerador y denominador, respectivamente.

Para el problema de dos muestras en consideración, la prueba F que acabamos de describir es equivalente a la prueba t de dos colas del Capítulo 10. Entonces, ¿por qué molestarse en establecer esta equivalencia? Como veremos en la Sección 13.3, la prueba F fácilmente se generaliza para permitir una comparación de cualquier número de tratamientos.

EJEMPLO **13.1** Los valores codificados para una medida de elasticidad de un plástico preparado por dos procesos diferentes se proporcionan en la Tabla 13.1. Las muestras independientes, ambas de tamaño 6, se tomaron de la producción de cada uno de los procesos. ¿Los datos presentan suficiente evidencia para indicar una diferencia en elasticidad media en los dos procesos?

Tabla 13.1
Datos para el Ejemplo 13.1

A	B
6.1	9.1
7.1	8.2
7.8	8.6
6.9	6.9
7.6	7.5
8.2	7.9

Solución Aunque la prueba t de dos muestras de la Sección 10.8 podría usarse para analizar estos datos, usaremos la prueba F ANOVA estudiada antes en esta sección. Las tres sumas de cuadrados buscadas son

$$\begin{aligned}
\text{SS total} &= \sum_{i=1}^{2}\sum_{j=1}^{6}(y_{ij}-\overline{y})^2 = \sum_{i=1}^{2}\sum_{j=1}^{6}y_{ij}^2 - \frac{1}{12}\left(\sum_{i=1}^{2}\sum_{j=1}^{6}y_{ij}\right)^2 \\
&= 711.35 - \frac{1}{12}(91.9)^2 = 7.5492, \\
\text{SST} &= n_1\sum_{i=1}^{2}(\overline{y}_i-\overline{y})^2 = 6\sum_{i=1}^{2}(\overline{y}_i-\overline{y})^2 = 1.6875, \\
\text{SSE} &= \sum_{i=1}^{2}\sum_{j=1}^{6}(y_{ij}-\overline{y}_i)^2 = 5.8617.
\end{aligned}$$

(Usted puede verificar que la SSE es la suma de cuadrados de las desviaciones agrupada para las dos muestras y que la SS total $=$ SST $+$ SSE.) Los cuadrados medios del tratamiento y el error son, respectivamente

$$\begin{aligned}
\text{MST} &= \frac{\text{SST}}{1} = 1.6875, \\
\text{MSE} &= \frac{\text{SSE}}{2n_1-2} = \frac{5.8617}{10} = .58617.
\end{aligned}$$

Para probar la hipótesis nula $\mu_1 = \mu_2$, calculamos el valor del estadístico de prueba

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{1.6875}{.58617} = 2.88$$

y rechazamos H_0 si el valor calculado de F excede a F_α. El valor crítico del estadístico F con 1 grado de libertad en el numerador y 10 grados de libertad en el denominador para $\alpha = .05$ y $F_{.05} = 4.96$. Aunque el cuadrado medio del tratamiento (MST) es casi tres veces el cuadrado medio del error (MSE), no es suficientemente grande para permitir el rechazo de la hipótesis nula. En consecuencia, en el nivel de significancia $\alpha = .05$, no hay suficiente evidencia para indicar una diferencia entre μ_1 y μ_2. El nivel de significancia alcanzado está dado por el valor $p = P(F > 2.88)$. De acuerdo con la Tabla 7, Apéndice 3, valor $p > .10$. La aplicación *F-Ratio Probabilities and Quantiles* proporciona el valor p exacto $= P(F > 2.88) = .12054$.

El propósito de este ejemplo es ilustrar los cálculos comprendidos en un ANOVA sencillo. La prueba F para comparar dos medias es equivalente a una prueba t de dos muestras porque el cuadrado de una variable aleatoria con distribución t con ν grados de libertad tiene una distribución F con 1 grado de libertad en el numerador y ν grados de libertad en el denominador. Usted puede verificar fácilmente que el cuadrado de $t_{.025} = 2.228$ (empleado para la prueba de dos colas con $\alpha = .05$ y $\nu = 10$ grados de libertad) es igual a $F_{.05} = 4.96$. Si se hubiera usado la prueba t para el Ejemplo 13.1, hubiéramos obtenido $t = -1.6967$, que satisface la relación $t^2 = (1.6967)^2 = 2.88 = F$. ■

Ejercicios

13.1 Los tiempos de reacción para dos estímulos diferentes en un experimento psicológico de asociación de palabras se compararon usando cada estímulo en muestras aleatorias independientes de tamaño 8. Así, un total de 16 personas se usaron en el experimento. ¿Los siguientes datos presentan suficiente evidencia para indicar que hay una diferencia en los tiempos medios de reacción para los dos estímulos?

Estímulo 1	1	3	2	1	2	1	3	2
Estímulo 2	4	2	3	3	1	2	3	3

a Use el método ANOVA para probar las hipótesis apropiadas. Prueba en el nivel $\alpha = .05$ de significancia.

b **Ejercicio Applet** Use la aplicaión *F-Ratio Probabilities and Quantiles* para determinar el valor p exacto para la prueba del inciso a.

c Pruebe las hipótesis apropiadas con el uso de la prueba t de dos muestras para comparar medias poblacionales que desarrollamos en la Sección 10.8. Compare el valor del estadístico t contra el valor del estadístico F calculado en el inciso a.

d ¿Qué suposiciones son necesarias para las pruebas realizadas en incisos anteriores?

13.2 Consulte los Ejercicios 8.90 y 10.77.

a Use una prueba F para determinar si hay suficiente evidencia para expresar una diferencia en las medias de las calificaciones SAT verbales, de estudiantes de secundaria que pretendan especializarse en ingeniería y lengua-literatura. Defina límites para el valor p asociado. ¿Qué concluiría usted con un nivel de significancia $\alpha = .05$?

b **Ejercicio Applet** Use la aplicación *F-Ratio Probabilities and Quantiles* para determinar el valor p exacto para la prueba del inciso a.

c ¿Cómo se compara el valor del estadístico F obtenido en el inciso a contra el valor del estadístico t obtenido en el Ejercicio 10.77?

d ¿Qué suposiciones son necesarias para los análisis efectuados en el inciso a?

13.3 Comparación de más de dos medias: análisis de varianza para un diseño de un factor

Un ANOVA para comparar más de dos medias poblacionales es una generalización del ANOVA presentado en la Sección 13.2. La selección aleatoria de muestras independientes de k poblaciones se conoce como *diseño de un factor*. Como se indica en la Sección 12.4, los datos en un diseño de un factor pueden corresponder a datos obtenidos de un diseño experimental completamente aleatorizado (véase la Definición 12.4) *o* de tomar muestras independientes de cada una de varias poblaciones existentes.

Suponga que muestras aleatorias independientes se han sacado de k poblaciones normales con medias $\mu_1, \mu_2, \ldots, \mu_k$, respectivamente, y varianza común σ^2. Para que sean completamente generales, permitiremos que los tamaños muestrales sean desiguales y que n_i, para $i = 1, 2, \ldots, k$ sea el número de observaciones de la muestra tomadas de la i-ésima población. El número total de observaciones del experimento es $n = n_1 + n_2 + \cdots + n_k$.

Denotemos con Y_{ij} la respuesta para la j-ésima unidad experimental de la i-ésima muestra y representemos con $Y_{i\bullet}$ y $\overline{Y}_{i\bullet}$ el total y la media, respectivamente, de las n_i respuestas en la i-ésima muestra. El punto en la segunda posición del subíndice de $Y_{i\bullet}$ tiene la finalidad de recordar al lector que esta cantidad se calcula sumando todos los posibles valores del subíndice que es sustituido por el punto, j en este caso. Del mismo modo, los subíndices de $\overline{Y}_{i\bullet}$ indican que esta media se calcula promediando los valores de la i-ésima muestra. Entonces, para $i = 1, 2, \ldots, k$,

$$Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij} \quad \text{y} \quad \overline{Y}_{i\bullet} = \left(\frac{1}{n_i}\right) \sum_{j=1}^{n_i} Y_{ij} = \left(\frac{1}{n_i}\right) Y_{i\bullet}.$$

Esta modificación de los símbolos para totales y promedios muestrales simplificará las fórmulas de cálculo para las sumas de cuadrados.

Entonces, como en el ANOVA que contiene dos medias, tenemos

$$\text{SS total} = \text{SST} + \text{SSE}$$

(demostración diferida a la Sección 13.6), donde

$$\text{SS total} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 - \text{CM},$$

$$\text{CM} = \frac{(\text{total de todas las observaciones})^2}{n} = \frac{1}{n}\left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}\right)^2 = n\overline{Y}^2,$$

(el símbolo CM denota *corrección para la media*),

$$\text{SST} = \sum_{i=1}^{k} n_i (\overline{Y}_{i\bullet} - \overline{Y})^2 = \sum_{i=1}^{k} \frac{Y_{i\bullet}^2}{n_i} - \text{CM},$$

$$\text{SSE} = \text{SS total} - \text{SST}.$$

Aunque la forma más fácil de calcular la SSE es por sustracción, como ya vimos antes, es interesante observar que la SSE es la suma agrupada de cuadrados para todas las k muestras y es igual a

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\bullet})^2 \\ &= \sum_{i=1}^{k} (n_i - 1) S_i^2, \end{aligned}$$

donde

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\bullet})^2.$$

Observe que la SSE es una función *sólo* de las varianzas muestrales S_i^2, para $i = 1, 2, \ldots, k$. Como cada uno de los S_i^2 valores proporciona un estimador insesgado para $\sigma_i^2 = \sigma^2$ con $n_i - 1$ grados de libertad, un estimador insesgado de σ^2 basado en $(n_1 + n_2 + \cdots + n_k - k) = n - k$ grados de libertad está dado por

$$S^2 = \text{MSE} = \frac{\text{SSE}}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1)} = \frac{\text{SSE}}{n - k}.$$

Como

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n} \sum_{i=1}^{k} n_i \overline{Y}_{i\bullet},$$

se deduce que SST es una función *sólo* de las medias muestrales $\overline{Y}_{i\bullet}$, para $i = 1, 2, \ldots, k$. El MST posee $(k - 1)$ grados de libertad, es decir, 1 menos que el número de medias, y es

$$\text{MST} = \frac{\text{SST}}{k - 1}.$$

Para probar la hipótesis nula,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k,$$

contra la alternativa de que al menos una de las igualdades no se cumple, comparamos MST con MSE, usando el estadístico F con base en $\nu_1 = k - 1$ y $\nu_2 = n - k$ grados de libertad en el numerador y el denominador, respectivamente. La hipótesis nula será rechazada si

$$F = \frac{\text{MST}}{\text{MSE}} > F_\alpha,$$

donde F_α es el valor crítico de F para una prueba de nivel α. En el Ejercicio 13.6 usted demostrará que dada $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, el estadístico F posee una distribución F con $k - 1$ y $n - k$ grados de libertad en el numerador y el denominador, respectivamente.

De acuerdo con las convenciones que establecimos, usaremos la notación y_{ij} para denotar el valor observado de Y_{ij}. Del mismo modo, usaremos $y_{i\bullet}$ y $\overline{y}_{i\bullet}$ para denotar los valores observados de $Y_{i\bullet}$ y $\overline{Y}_{i\bullet}$,para $i = 1, 2, \ldots, k$, respectivamente. De manera intuitiva, cuanto mayores sean las diferencias entre los valores observados de las medias de tratamiento, $\overline{y}_{1\bullet}, \overline{y}_{2\bullet}, \ldots, \overline{y}_{k\bullet}$, mayor es la evidencia para indicar una diferencia entre las medias poblacionales correspondientes. Si todas las medias de tratamiento son idénticas, $\overline{y}_{1\bullet} = \overline{y}_{2\bullet} = \cdots = \overline{y}_{k\bullet} = \overline{y}$ y todas las diferencias que aparecen en la expresión anterior para SST son iguales a cero, implica que SST = 0. Conforme las medias para tratamiento se alejan entre sí, las desviaciones $(\overline{y}_{i\bullet} - \overline{y})$ aumentan en valor absoluto y el valor observado de la SST aumenta en magnitud. En consecuencia, cuanto mayor sea el valor observado de la SST, mayor es el peso de la evidencia que hay a favor de rechazar la hipótesis nula. Esta misma línea de razonamiento se aplica a las pruebas F empleadas en el ANOVA para todos los experimentos diseñados.

Las suposiciones que sirven de base a las pruebas F del ANOVA merecen particular atención. Se supone que las muestras aleatorias independientes han sido seleccionadas de las k poblaciones. Se supone que las k poblaciones están distribuidas normalmente con varianzas $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ y medias $\mu_1, \mu_2, \ldots, \mu_k$. Las desviaciones moderadas respecto de estas suposiciones no afectarán de manera grave las propiedades de la prueba. Esto es particularmente cierto en la suposición de normalidad. La suposición de varianzas poblacionales iguales es menos crítica si los tamaños de las muestras de las poblaciones respectivas son todos iguales ($n_1 = n_2 = \cdots = n_k$). Se dice que un diseño de un factor con igual número de observaciones por tratamiento está *balanceado*.

EJEMPLO **13.2** Cuatro grupos de estudiantes se someten a diferentes técnicas de enseñanza y se examinan al final de un periodo especificado. Como consecuencia de las deserciones de los grupos experimentales (por enfermedad, transferencia, etc.), el número de estudiantes varió de un grupo a otro. ¿Los datos mostrados en la Tabla 13.2 presentan suficiente evidencia para indicar una diferencia en el éxito medio para las cuatro técnicas de enseñanza?

Solución Los valores observados de las cantidades necesarias para calcular el valor del estadístico F son

$$\text{CM} = \frac{1}{n}\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}\right)^2 = \frac{(1779)^2}{23} = 137\ 601.8,$$

$$\text{SS total} = \sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 - \text{CM} = 139\ 511 - 137\ 601.8 = 1909.2,$$

$$\text{SST} = \sum_{i=1}^{4}\frac{y_{i\bullet}^2}{n_i} - \text{CM} = 138\ 314.4 - 137\ 601.8 = 712.6,$$

$$\text{SSE} = \text{SS total} - \text{SST} = 1196.6.$$

Los valores observados de MST y MSE son

$$\text{MST} = \frac{\text{SST}}{k-1} = \frac{712.6}{3} = 237.5,$$

$$\text{MSE} = \frac{\text{SSE}}{n-k} = \frac{1196.6}{19} = 63.0.$$

Tabla 13.2 Datos para el Ejemplo 13.2

	1	2	3	4
	65	75	59	94
	87	69	78	89
	73	83	67	80
	79	81	62	88
	81	72	83	
	69	79	76	
		90		
$y_{i\bullet}$	454	549	425	351
n_i	6	7	6	4
$\overline{y}_{i\bullet}$	75.67	78.43	70.83	87.75

Finalmente, el valor observado del estadístico de prueba para probar la hipótesis nula $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ es

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{237.5}{63.0} = 3.77,$$

donde los grados de libertad apropiados en el numerador y el denominador son $\nu_1 = k - 1 = 3$ y $\nu_2 = n - k = (6 + 7 + 6 + 4) - 4 = 19$, respectivamente.

El nivel de significancia alcanzado está dado por valor $p = P(F > 3.77)$. Usando la Tabla 7, Apéndice 3, con 3 grados de libertad en el numerador y 19 en el denominador, vemos que $.025 <$ valor $p < .05$. Entonces, si escogemos $\alpha = .05$ (o cualquier valor mayor), rechazamos la hipótesis nula y concluimos que hay suficiente evidencia para indicar una diferencia en el éxito medio entre los cuatro procedimientos de enseñanza. La aplicación *F-Ratio Probabilities and Quantiles* se puede usar para establecer el valor p exacto $= P(F > 3.77) = .02808$. ■

Usted puede pensar que esta conclusión pudo hacerse con base en la observación visual de las medias de tratamiento. Sin embargo, no es difícil construir un conjunto de datos que lleve a resultados erróneos a quien tome decisiones visuales.

13.4 Tabla de análisis de varianza para un diseño de un factor

Los cálculos para un ANOVA suelen mostrarse en una tabla ANOVA (o AOV). La tabla para el diseño en la Sección 13.3 para comparar k medias de tratamiento se ilustra en la Tabla 13.3. La primera columna muestra la fuente asociada con cada una de las sumas de cuadrados; la segunda columna da los grados de libertad respectivos; las columnas tercera y cuarta dan las sumas de cuadrados y cuadráticas medias, respectivamente. Un valor calculado de F, comparando MST y MSE, suele aparecer en la quinta columna. Observe que SST + SSE = SS total y que la suma de grados de libertad del tratamiento y del error es igual al número total de grados de libertad.

Tabla 13.3 Tabla ANOVA para diseño de un factor

Fuente	gl	SS	MS	F
Tratamientos	$k-1$	SST	$\text{MST} = \dfrac{\text{SST}}{k-1}$	$\dfrac{\text{MST}}{\text{MSE}}$
Error	$n-k$	SSE	$\text{MSE} = \dfrac{\text{SSE}}{n-k}$	
Total	$n-1$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2$		

Tabla 13.4 Tabla ANOVA para el Ejemplo 13.2

Fuente	gl	SS	MS	F
Tratamientos	3	712.6	237.5	3.77
Error	19	1196.6	63.0	
Total	22	1909.2		

La tabla ANOVA para el Ejemplo 13.2, que se muestra en la Tabla 13.4, constituye una forma de presentación compacta de las cantidades calculadas apropiadas para el análisis de varianza.

Ejercicios

13.3 Enuncie las suposiciones que sirven de base al ANOVA de un diseño completamente aleatorizado.

13.4 Consulte el Ejemplo 13.2. Calcule el valor de la SSE al agrupar las sumas de los cuadrados de las desviaciones dentro de cada una de las cuatro muestras y compare la respuesta con el valor obtenido por sustracción. Ésta es una extensión del procedimiento de agrupación empleado en el caso de dos muestras estudiado en la Sección 13.2.

***13.5** En el Ejercicio 6.59 mostramos que si Y_1 y Y_2 son variables aleatorias independientes con distribución χ^2, con ν_1 y ν_2 grados de libertad, respectivamente, entonces $Y_1 + Y_2$ tienen una distribución χ^2 con $\nu_1 + \nu_2$ grados de libertad. Ahora suponga que $W = U + V$, donde U y V son variables aleatorias *independientes*, y que W y V tienen distribuciones χ^2 con r y s grados de libertad, respectivamente, donde $r > s$. Use el método de funciones generadoras de momentos para demostrar que U debe tener una distribución χ^2 con $r - s$ grados de libertad.[1]

13.6 Suponga que muestras independientes de tamaños $n_1, n_2, \ldots, n_k$ se toman de cada una de las k poblaciones distribuidas normalmente con medias $\mu_1, \mu_2, \ldots, \mu_k$ y varianzas comunes, todas iguales a σ^2. Denotemos con Y_{ij} la j-ésima observación de la población i, para $j = 1, 2, \ldots, n_i$ e $i = 1, 2, \ldots, k$, y sea $n = n_1 + n_2 + \cdots + n_k$.

a Recuerde que

$$\text{SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2 \qquad \text{donde } S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2.$$

Demuestre que SSE/σ^2 tiene una distribución χ^2 con $(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n - k$ grados de libertad.

1. Los ejercicios precedidos por un asterisco son opcionales.

b Demuestre que de acuerdo con la hipótesis nula, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ todas las Y_{ij} son variables aleatorias independientes y distribuidas normalmente con la *misma media y varianza.* Use el Teorema 7.3 para demostrar además que, dada la hipótesis nula,

$$\text{SS Total} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2$$

es tal que (SS total)$/\sigma^2$ tiene una distribución χ^2 con $n - 1$ grados de libertad.

c En la Sección 13.3 dijimos que la SST es una función sólo de las medias muestrales y que la SSE es una función sólo de las varianzas muestrales. En consecuencia, las SST y SSE son independientes. Recuerde que SST total = SST + SSE. Use los resultados del Ejercicio 13.5 y los incisos a y b para demostrar que, dada la hipótesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, SST/$\sigma^2$ tiene una distribución χ^2 con $k - 1$ grados de libertad.

d Use los resultados de los incisos a–c para afirmar que, de acuerdo con la hipótesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, $F = \text{MST}/\text{MSE}$ tiene una distribución F con $k - 1$ y $n - k$ grados de libertad en el numerador y el denominador, respectivamente.

13.7 Cuatro plantas químicas, que producen los mismos productos y son propiedad de la misma compañía, descargan aguas residuales en arroyos en la cercanía donde se encuentran ubicadas. Para vigilar la magnitud de la contaminación creada por las aguas residuales y para determinar si esto difiere de una planta a otra, la compañía recolectó aleatoriamente muestras de desechos líquidos, cinco muestras de cada planta. Los datos se presentan en la siguiente tabla.

Planta	Aguas residuales contaminantes (lb/gal de desechos)				
A	1.65	1.72	1.50	1.37	1.60
B	1.70	1.85	1.46	2.05	1.80
C	1.40	1.75	1.38	1.65	1.55
D	2.10	1.95	1.65	1.88	2.00

a ¿La información aporta suficiente evidencia para indicar una diferencia en el peso medio de las aguas residuales por galón descargadas de las cuatro plantas? Pruebe usando $\alpha = .05$.

b **Ejercicio Applet** Encuentre el valor p asociado con la prueba del inciso a usando la aplicación *F-Ratio Probabilities and Quantiles.*

13.8 En un estudio de salarios iniciales para profesores auxiliares, cinco profesores auxiliares hombres de tres tipos de instituciones que otorgan doctorados se encuestaron aleatoriamente; se registraron sus salarios iniciales pero sus nombres se mantuvieron en el anonimato. Los resultados de la encuesta (medidos en miles de dólares) se proporcionan en la tabla siguiente.[2]

Universidades públicas	Privadas-independientes	Afiliadas a la iglesia
49.3	81.8	66.9
49.9	71.2	57.3
48.5	62.9	57.7
68.5	69.0	46.2
54.0	69.0	52.2

2. *Fuente:* adaptado de "Average Salary for Men and Women Faculty, by Category, Affiliation, and Academy Rank 2002–2003," *Academe: Bulletin of the American Association of University Professors*, marzo-abril de 2003, 37.

a ¿Qué tipo de diseño experimental se utilizó cuando se recolectaron los datos?

b ¿Hay suficiente evidencia para indicar una diferencia en el promedio de los salarios iniciales de profesores auxiliares en los tres tipos de instituciones que otorgan doctorados? Use la tabla de la prueba para limitar el valor p.

c **Ejercicio Applet** Determine el valor p exacto con el uso de la aplicación *F-Ratio Probabilities and Quantiles.*

13.9 En una comparación de las resistencias del concreto producido con cuatro mezclas se prepararon tres muestras de cada tipo de mezcla. Cada una de las 12 muestras fue sometida a cargas de compresión cada vez mayores hasta romperlas. La siguiente tabla proporciona las cargas de compresión, en toneladas por pulgada cuadrada, alcanzadas en el punto de ruptura. Los números de muestra 1–12 están indicados entre paréntesis para fines de identificación.

Mezcla A	Mezcla B	Mezcla C	Mezcla D
(1) 2.30	(2) 2.20	(3) 2.15	(4) 2.25
(5) 2.20	(6) 2.10	(7) 2.15	(8) 2.15
(9) 2.25	(10) 2.20	(11) 2.20	(12) 2.25

a Analice la información suponiendo que se satisfacen las necesidades para un diseño de un factor. Diga si hay apoyo estadístico con un nivel de significancia $\alpha = .05$ para la conclusión de que al menos uno de los concretos difiere en resistencia promedio con respecto a los otros.

b **Ejercicio Applet** Use la aplicación *F-Ratio Probabilities and Quantiles* para hallar el valor p asociado con la prueba del inciso a.

13.10 Un psicólogo clínico deseaba comparar tres métodos para reducir los niveles de hostilidad en estudiantes universitarios. Se utilizó un examen psicológico (HLT) para medir el grado de hostilidad. Puntuaciones altas en este examen indican gran hostilidad; en el experimento se emplearon once estudiantes que obtienen calificaciones altas y casi iguales. Cinco se tomaron aleatoriamente de entre los 11 casos problema y fueron tratados por el método A. Tres se tomaron aleatoriamente de los 6 estudiantes restantes y fueron tratados por el método B. Los otros 3 estudiantes fueron tratados por el método C. Todos los tratamientos continuaron durante un semestre. A cada estudiante se le aplicó un examen HLT de nuevo al finalizar el semestre con los resultados mostrados en la siguiente tabla.

Método A	Método B	Método C
73	54	79
83	74	95
76	71	87
68		
80		

a ¿Los datos aportan suficiente evidencia para indicar que al menos uno de los métodos de tratamiento produce una respuesta media del estudiante diferente de los otros métodos? Fije límites para el nivel de significancia alcanzado.

b **Ejercicio Applet** Determine el valor p exacto con el uso de la aplicación *F-Ratio Probabilities and Quantiles.*

c ¿Qué concluiría usted en el nivel de significancia $\alpha = .05$?

13.11 Se cree que las mujeres en la fase posmenopáusica de su vida sufren de deficiencia de calcio. Este fenómeno está asociado con la proporción relativamente alta de fracturas de hueso en mujeres de ese grupo de edad. ¿La deficiencia de calcio está asociada con la deficiencia de estrógeno, afección que se presenta después de la menopausia? Para investigar esta teoría, L. S. Richelson y colegas[3] compararon la densidad mineral en huesos de tres grupos de mujeres.

El primer grupo de 14 mujeres había padecido de ooforectomía (extirpación quirúrgica de ovarios) durante su edad adulta joven y había vivido un periodo de 15 a 25 años con deficiencia de estrógeno. Un segundo grupo, identificado como premenopáusico, tenía más o menos la misma edad (aproximadamente 50 años) que el grupo de ooforectomía, excepto que las mujeres nunca habían sufrido un periodo de deficiencia de estrógeno. El tercer grupo de 14 mujeres eran posmenopáusicas y habían sufrido de deficiencia de estrógenos durante un promedio de 20 años. La media y el error estándar de la media para las tres muestras de mediciones de densidad ósea en la espina dorsal, es decir, 14 mediciones en cada muestra, una por cada persona, se registraron en la siguiente tabla.

Ooforectomía grupo 1		Premenopáusicas grupo 2		Posmenopáusicas grupo 3	
Media	Error estándar	Media	Error estándar	Media	Error estándar
0.93	0.04	1.21	0.03	0.92	0.04

a ¿Hay suficiente evidencia que permita concluir que las mediciones medias de densidad ósea difieren para los tres grupos de mujeres? ¿Cuál es el valor p asociado con su prueba?

b ¿Qué concluiría usted con un nivel de significancia $\alpha = .05$?

13.12 Si las legumbres destinadas al consumo humano contienen algún plaguicida, éste debe presentarse en cantidades pequeñísimas. La detección de plaguicidas en legumbres enviadas al mercado se logra mediante solventes para extraer los plaguicidas de las legumbres y luego realizar pruebas en estos extractos para aislar y cuantificar los plaguicidas presentes. Se cree que el proceso de extracción es adecuado porque, si se agregan cantidades conocidas de plaguicida a legumbres "limpias" en un ambiente de laboratorio, en esencia todos los plaguicidas se pueden eliminar del extracto contaminado de manera artificial.

Los siguientes datos se obtuvieron de un estudio realizado por Willis Wheeler y colegas,[4] que buscaron determinar si el proceso de extracción también es eficaz cuando se usa en situaciones más reales donde se aplican plaguicidas a cosechas de legumbres. El Dieldrin (plaguicida de uso común) marcado con carbono-14 (radiactivo) se aplicó a rábanos en crecimiento. Catorce días después se empleó el proceso de extracción y se analizaron los extractos en busca de algún contenido de plaguicida. Un contador de centelleo en líquido se utilizó para determinar la cantidad de carbono-14 presente en el extracto y también la cantidad dejada en la pulpa de la legumbre. Como por lo general la pulpa de la legumbre se desecha cuando es analizada en busca de plaguicidas, si una proporción apreciable de plaguicida continúa en la pulpa podría resultar en una subestimación de la cantidad de plaguicida. Éste fue la única fuente de carbono-14 y, por tanto, es probable que la proporción de carbono-14 en la pulpa sea indicativa de la proporción de plaguicida en la pulpa. La siguiente tabla muestra una parte de los datos que los investigadores obtuvieron cuando se emplearon concentraciones bajas, medias y altas del solvente, acetonitrilo, en el proceso de extracción.

3. *Fuente:* L. S. Richelson, H. W. Wahner, L. J. Melton III, and B. L. Riggs, "Relative Contributions of Aging and Estrogen Deficiency to Postmenopausal Bone Loss," *New England Journal of Medicine* 311(20) (1984): 1273-1275.

4. *Fuente:* Willis B. Wheeler, N. P. Thompson, R. L. Edelstein, R. C. Littel, and R. T. Krause, "Influence of Various Solvent-Water Mixtures on the Extraction of Dieldrin and Methomyl Residues from Radishes," *Journal of the Association of Official Analytical Chemists* 65(5) (1982): 1112–1117.

Porcentaje de carbono-14 en pulpa de legumbre			
	Concentración de acetonitrilo		
	Baja	Media	Alta
	23.37	20.39	18.87
	25.13	20.87	19.69
	23.78	20.78	19.29
	27.74	20.19	18.10
	25.30	20.01	18.42
	25.21	20.23	19.33
	22.12	20.73	17.26
	20.96	19.53	18.09
	23.11	18.87	18.69
	22.57	18.17	18.82
	24.59	23.34	18.72
	23.70	22.45	18.75
Total	287.58	245.56	224.03

a ¿Hay suficiente evidencia de que el porcentaje medio de carbono-14 restante en la pulpa de la legumbre difiere para las diferentes concentraciones de acetonitrilo usadas en el proceso de extracción? Fije límites o use la aplicación apropiada para determinar el nivel de significancia alcanzado. ¿Qué concluiría con un nivel de significancia de $\alpha = .01$?

b ¿Qué suposiciones son necesarias para emplear de manera válida el análisis realizado en el inciso a? Relacione las suposiciones necesarias con la aplicación específica representada en este ejercicio.

13.13 Una parte de la investigación descrita en un artículo científico de Yean-Jye Lu[5] comprendía una evaluación de tiempos de maniobra para vehículos de varios tamaños que daban vuelta a la izquierda en un crucero con un carril propio de vuelta a la izquierda pero sin una señal independiente de vuelta en esta dirección, en el semáforo que controlaba el crucero (una maniobra "desprotegida" de vuelta a la izquierda). Se midió el tiempo de maniobra desde el instante en que un vehículo entraba a los carriles opuestos al tránsito, hasta que pasaba por completo por el crucero. Autos de cuatro cilindros fueron clasificados como "pequeños" y los de seis y ocho cilindros como "grandes". Camiones y autobuses se combinaron para formar una tercera categoría identificada como "camión o autobús". Otros vehículos motorizados (motocicletas, etc.) se ignoraron en el estudio. Un resumen de los datos, que contiene los tiempos de maniobra (en segundos) para vehículos que intentaban la maniobra de dar vuelta a la izquierda desde una posición de alto total, aparece en la siguiente tabla.

Tipo de vehículo	Tamaño muestral	Media	Desviación estándar
Auto pequeño	45	4.59	0.70
Auto grande	102	4.88	0.64
Camión o autobús	18	6.24	0.90

a ¿Hay suficiente evidencia para decir que los tiempos medios de maniobra difieren para los tres tipos de vehículo? Establezca límites para el nivel de significancia alcanzado.

b Indique la conclusión apropiada para una prueba de nivel $\alpha = .05$.

5. Fuente: Yean-Jye Lu, "A Study of Left-Turn Maneuver Time for Signalized Intersections," *ITE Journal* 54 (octubre de 1984): 42–47. Institute of Transportation Engineers, Washington, D.C., © 1984 I.T.E. Todos los derechos reservados.

13.14 La Comisión de Caza y Pesca de la Florida desea comparar las cantidades de residuos de tres sustancias químicas halladas en el tejido cerebral de pelícanos cafés. Muestras aleatorias independientes de diez pelícanos aportaron cada una los siguientes resultados (medidos en partes por millón). ¿Hay evidencia de diferencias importantes entre las cantidades medias de residuos, con un nivel de significancia de 5%?

	Sustancia química		
Estadístico	DDE	DDD	DDT
Media	.032	.022	.041
Desviación estándar	.014	.008	.017

13.15 Se tomaron muestras de agua en cuatro lugares diferentes en un río para determinar si la cantidad de oxígeno disuelto, que es una medida de la contaminación del agua, difería de un lugar a otro. Los lugares 1 y 2 se seleccionaron aguas arriba de una planta industrial, uno cerca de la orilla y el otro a media corriente; el lugar 3 estaba adyacente a la descarga de aguas industriales de la planta; y el lugar 4 estaba ligeramente aguas abajo a media corriente. Se seleccionaron aleatoriamente cinco muestras de agua en cada lugar, pero una de ellas, la del lugar 4, se perdió en el laboratorio. Los datos se muestran en la siguiente tabla (cuanto mayor es la contaminación, menores serán las lecturas de oxígeno disuelto). ¿Los datos arrojan suficiente evidencia para indicar una diferencia en el contenido medio de oxígeno disuelto para los cuatro lugares? Establezca límites para el nivel de significancia alcanzado.

Lugar	Contenido de oxígeno disuelto				
1	5.9	6.1	6.3	6.1	6.0
2	6.3	6.6	6.4	6.4	6.5
3	4.8	4.3	5.0	4.7	5.1
4	6.0	6.2	6.1	5.8	

13.16 Se realizó un experimento para examinar el efecto de la edad en la frecuencia cardiaca cuando las personas hacían una cantidad específica de ejercicio. Diez hombres se seleccionaron aleatoriamente de cuatro grupos de edad: 10–19, 20–39, 40–59 y 60–69. Cada persona hizo ejercicio en una caminadora, a un ritmo fijo durante 12 minutos y se le registró el aumento en la frecuencia cardiaca (en pulsaciones por minuto), es decir, la diferencia en frecuencia antes y después del ejercicio. Cálculos preliminares dieron SS total $= 1002.975$ y SST $= 67.475$.

a Construya la tabla ANOVA asociada.

b ¿Los datos aportan suficiente evidencia para indicar diferencias en el aumento medio en la frecuencia cardiaca entre los cuatro grupos de edad? Efectúe la prueba usando $\alpha = .05$.

13.5 Modelo estadístico para el diseño de un factor

Igual que antes, denotemos con Y_{ij} las variables aleatorias que generan los valores observados y_{ij}, para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$. Los valores Y_{ij} corresponden a muestras aleatorias independientes de poblaciones normales con $E(Y_{ij}) = \mu_i$ y $V(Y_{ij}) = \sigma^2$, para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$. Consideremos la muestra aleatoria tomada de la población 1 y escribamos

$$Y_{1j} = \mu_1 + \varepsilon_{1j}, \qquad j = 1, 2, \ldots, n_1.$$

De modo equivalente,

$$\varepsilon_{1j} = Y_{1j} - \mu_1, \qquad j = 1, 2, \ldots, n_1.$$

Como ε_{1j} es la diferencia entre una variable aleatoria distribuida normalmente y su media, se deduce que ε_{1j} está distribuida normalmente con $E(\varepsilon_{1j}) = 0$ y $V(\varepsilon_{1j}) = V(Y_{1j}) = \sigma^2$. Además, la independencia de Y_{1j}, para $j = 1, 2, \ldots, n_1$, implica que $\varepsilon_{1j} = 1, 2, \ldots, n_1$, son variables aleatorias mutuamente independientes. Para cada $i = 1, 2, \ldots, k$, podemos proceder de modo análogo para escribir

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad j = 1, 2, \ldots, n_i,$$

donde los "términos de error" ε_{ij} son variables aleatorias independientes, distribuidas normalmente, con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$, para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$. Los términos de error simplemente representan la diferencia entre las observaciones de cada muestra y las medias poblacionales correspondientes.

Un conjunto más de consideraciones llevará al modelo clásico para el diseño de un factor. Considere las medias μ_i, para $i = 1, 2, \ldots, k$, y escriba

$$\mu_i = \mu + \tau_i \qquad \text{donde } \tau_1 + \tau_2 + \cdots + \tau_k = 0.$$

Observe que $\sum_{i=1}^{k} \mu_i = k\mu + \sum_{i=1}^{k} \tau_i = k\mu$ y por tanto $\mu = k^{-1} \sum_{i=1}^{k} \mu_i$ es sólo el promedio de las k medias poblacionales (los valores μ_i). Por esta razón, μ suele citarse como la *media total*. Como para $i = 1, 2, \ldots, k$, $\tau_i = \mu_i - \mu$ cuantifica la diferencia entre la media para la población i y la media total, τ_i normalmente recibe el nombre de *efecto de tratamiento* (o población) i. Por último, presentamos el modelo clásico para el diseño de un factor.

Modelo estadístico para un diseño de un factor

Para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$,

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

donde Y_{ij} = la j-ésima observación de la población (tratamiento) i,
μ = la media general,
τ_i = el efecto no aleatorio del tratamiento i, donde $\sum_{i=1}^{k} \tau_i = 0$,
ε_{ij} = términos de error aleatorio tales que ε_{ij} son variables aleatorias independientes, distribuidas normalmente, con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$.

La ventaja de este modelo es que resume con gran claridad todas las suposiciones hechas en el análisis de los datos obtenidos de un diseño de un factor. También nos da una base para presentar un modelo estadístico preciso para el diseño de bloques aleatorizado. (Vea la Sección 13.8.)

Observe que (véase el Ejercicio 13.19) $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ se puede expresar también como

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_k = 0$$

y que $H_a : \mu_i \neq \mu_{i'}$ para alguna $i \neq i'$ es equivalente a $H_a : \tau_i \neq 0$ para alguna i, $1 \leq i \leq k$. Entonces, la prueba F para igualdad de medias que presentamos en la Sección 13.3 es la prueba de las hipótesis

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0 \quad \text{contra } H_a : \tau_i \neq 0 \text{ para alguna } i,\ 1 \leq i \leq k.$$

Ejercicios

13.17 Con $\overline{Y}_{i\bullet}$ denote el promedio de todas las respuestas al tratamiento i. Use el modelo para el diseño de un factor para deducir $E(\overline{Y}_{i\bullet})$ y $V(\overline{Y}_{i\bullet})$.

13.18 Consulte el Ejercicio 13.17 y considere $\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ para $i \neq i'$.

a Demuestre que $E(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = \mu_i - \mu_{i'} = \tau_i - \tau_{i'}$. Este resultado implica que $\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ es un estimador insesgado de la diferencia en los efectos de tratamientos i e i'.

b Deduzca $V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet})$.

13.19 Consulte el modelo estadístico para el diseño de un factor.

a Demuestre que $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$ es equivalente a $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$.

b Demuestre que $H_a : \tau_i \neq 0$ para al menos una i es equivalente a $H_a : \mu_i \neq \mu_{i'}$ para alguna $i \neq i'$.

13.6 Prueba de aditividad de las sumas de cuadrados y E(MST) para un diseño de un factor (opcional)

La demostración de que

$$\text{SS total} = \text{SST} + \text{SSE}$$

para el diseño de un factor se presenta en esta sección para beneficio de quienes estén interesados. Puede omitirse sin pérdida de continuidad.

La demostración usa resultados elementales en sumatorias que aparecen en los ejercicios del Capítulo 1 y la técnica de sumar y restar $\overline{Y}_{i\bullet}$ dentro de la expresión para el SS total. Entonces,

$$\begin{aligned}
\text{SS total} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet} + \overline{Y}_{i\bullet} - \overline{Y})^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(Y_{ij} - \overline{Y}_{i\bullet}) + (\overline{Y}_{i\bullet} - \overline{Y})]^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(Y_{ij} - \overline{Y}_{i\bullet})^2 + 2(Y_{ij} - \overline{Y}_{i\bullet})(\overline{Y}_{i\bullet} - \overline{Y}) + (\overline{Y}_{i\bullet} - \overline{Y})^2].
\end{aligned}$$

Sumando primero para j, obtenemos

$$\text{SS total } = \sum_{i=1}^{k}\left[\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2 + 2(\overline{Y}_{i\bullet} - \overline{Y})\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet}) + n_i(\overline{Y}_{i\bullet} - \overline{Y})^2\right],$$

donde

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet}) = Y_{i\bullet} - n_i\overline{Y}_{i\bullet} = Y_{i\bullet} - Y_{i\bullet} = 0.$$

En consecuencia, el término medio de la expresión para el SS total es igual a cero.

Entonces, sumando para i, obtenemos

$$\text{SS total } = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2 + \sum_{i=1}^{k} n_i(\overline{Y}_{i\bullet} - \overline{Y})^2 = \text{SSE} + \text{SST}.$$

La demostración de aditividad de las sumas de cuadrados ANOVA para otros diseños experimentales se puede obtener de un modo semejante, aunque el procedimiento es a veces tedioso.

A continuación proseguimos con la deducción del valor esperado del MST para un diseño de un factor (incluyendo un diseño completamente aleatorizado). Usando el modelo estadístico para el diseño de un factor presentado en la Sección 13.5 se deduce que

$$\overline{Y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i}\sum_{j=1}^{n_i}(\mu + \tau_i + \varepsilon_{ij}) = \mu + \tau_i + \overline{\varepsilon}_i, \qquad \text{donde } \overline{\varepsilon}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij}.$$

Como las ε_{ij} son variables aleatorias independientes con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$, el Teorema 5.12 implica (vea el Ejemplo 5.27) que $E(\overline{\varepsilon}_i) = 0$ y $V(\overline{\varepsilon}_i) = \sigma^2/n_i$.

De un modo análogo, $\overline{Y}$ está dada por

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mu + \tau_i + \varepsilon_{ij}) = \mu + \overline{\tau} + \overline{\varepsilon},$$

donde

$$\overline{\tau} = \frac{1}{n}\sum_{i=1}^{k} n_i\tau_i \quad \text{y} \quad \overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\varepsilon_{ij}.$$

Como los valores τ_i son constantes, $\overline{\tau}$ es simplemente una constante; de nuevo, usando el Teorema 5.12, obtenemos $E(\overline{\varepsilon}) = 0$ y $V(\overline{\varepsilon}) = \sigma^2/n$.

Por tanto, con respecto a los términos del modelo para el diseño de un factor,

$$\begin{aligned}
\text{MST} &= \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\overline{Y}_{i\bullet} - \overline{Y})^2 = \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\tau_i + \overline{\varepsilon}_i - \overline{\tau} - \overline{\varepsilon})^2 \\
&= \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\tau_i - \overline{\tau})^2 + \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} 2n_i(\tau_i - \overline{\tau})(\overline{\varepsilon}_i - \overline{\varepsilon}) \\
&\quad + \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\overline{\varepsilon}_i - \overline{\varepsilon})^2.
\end{aligned}$$

Como $\overline{\tau}$ y τ_i, para $i = 1, 2, \ldots, k$, son constantes y $E(\varepsilon_{ij}) = E(\overline{\varepsilon}_i) = E(\overline{\varepsilon}) = 0$, se deduce que

$$E(\text{MST}) = \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\tau_i - \overline{\tau})^2 + \left(\frac{1}{k-1}\right) E\left[\sum_{i=1}^{k} n_i(\overline{\varepsilon}_i - \overline{\varepsilon})^2\right].$$

Observe que

$$\begin{aligned}\sum_{i=1}^{k} n_i(\overline{\varepsilon}_i - \overline{\varepsilon})^2 &= \sum_{i=1}^{k}\left(n_i\overline{\varepsilon}_i^2 - 2n_i\overline{\varepsilon}_i\overline{\varepsilon} + n_i\overline{\varepsilon}^2\right)\\ &= \sum_{i=1}^{k} n_i\overline{\varepsilon}_i^2 - 2n\overline{\varepsilon}^2 + n\overline{\varepsilon}^2 = \sum_{i=1}^{k} n_i\overline{\varepsilon}_i^2 - n\overline{\varepsilon}^2.\end{aligned}$$

Como $E(\overline{\varepsilon}_i) = 0$ y $V(\overline{\varepsilon}_i) = \sigma^2/n_i$, se deduce que $E(\overline{\varepsilon}_i^2) = \sigma^2/n_i$, para $i = 1, 2, \ldots, k$. Del mismo modo, $E(\overline{\varepsilon}^2) = \sigma^2/n$ y, por tanto,

$$E\left[\sum_{i=1}^{k} n_i(\overline{\varepsilon}_i - \overline{\varepsilon})^2\right] = \sum_{i=1}^{k} n_i E\left(\overline{\varepsilon}_i^2\right) - nE\left(\overline{\varepsilon}^2\right) = k\sigma^2 - \sigma^2 = (k-1)\sigma^2.$$

Resumiendo, obtenemos

$$E(\text{MST}) = \sigma^2 + \left(\frac{1}{k-1}\right)\sum_{i=1}^{k} n_i(\tau_i - \overline{\tau})^2, \qquad \text{donde } \overline{\tau} = \frac{1}{n}\sum_{i=1}^{k} n_i\tau_i.$$

Dada $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$, se deduce que $\overline{\tau} = 0$ y, por tanto, $E(\text{MST}) = \sigma^2$. Entonces, cuando H_0 es verdadera, MST/MSE es la proporción entre dos estimadores insesgados para σ^2. Cuando $H_a : \tau_i \neq 0$ para alguna i, $1 \leq i \leq k$ es verdadera, la cantidad $1/(k-1)\sum_{i=1}^{k} n_i(\tau_i - \overline{\tau})^2$ es estrictamente positiva y MST es un estimador positivamente sesgado para σ^2.

13.7 Estimación en un diseño de un factor

Los intervalos de confianza para una media de tratamiento y para la diferencia entre un par de medias de tratamiento, basados en datos obtenidos en un diseño de un factor (Sección 13.3) son por completo análogos a los dados en el Capítulo 8. La única diferencia entre los intervalos del Capítulo 8 y los que siguen es que los intervalos asociados con el diseño de un factor usan el MSE (el estimador agrupado basado en todas las k muestras) para estimar la(s) varianza(s) poblacional(es) σ^2. El intervalo de confianza para la media de tratamiento i o la diferencia entre las medias para tratamientos i e i' son, respectivamente, como sigue:

$$\overline{Y}_{i\bullet} \pm t_{\alpha/2}\frac{S}{\sqrt{n_i}},$$

y

$$(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) \pm t_{\alpha/2}S\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}},$$

donde

$$S = \sqrt{S^2} = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n_1 + n_2 + \cdots + n_k - k}}$$

y $t_{\alpha/2}$ está basada en $(n - k)$ grados de libertad.

Los intervalos de confianza que acabamos de expresar son apropiados para una media de tratamiento o una comparación de un par de medias seleccionadas antes de la observación de los datos. Es probable que estos intervalos sean más cortos que los intervalos correspondientes del Capítulo 8 porque el valor de $t_{\alpha/2}$ está basado en un mayor número de grados de libertad ($n - k$ en lugar de $n_i - 1$ o $n_i + n_{i'} - 2$, respectivamente). Los coeficientes de confianza expresados son apropiados para una media o diferencia en dos medias identificadas *antes de observar los datos reales.* Si observáramos los datos y siempre comparáramos las poblaciones que produjeron la máxima y mínima medias muestrales, esperaríamos que la diferencia entre estas medias muestrales fuera más grande que para un par de medias especificado como de interés antes de observar los datos.

EJEMPLO **13.3** Encuentre un intervalo de confianza de 95% para la calificación media para la técnica de enseñanza 1, Ejemplo 13.2.

Solución El intervalo de confianza de 95% para la calificación media es

$$\overline{Y}_{1\bullet} \pm t_{.025}\frac{S}{\sqrt{n_1}},$$

donde $t_{.025}$ está determinada por $n - k = 19$ grados de libertad o

$$75.67 \pm (2.093)\frac{\sqrt{63}}{\sqrt{6}} \quad \text{o} \quad 75.67 \pm 6.78.$$

Observe que si hubiéramos analizado sólo los datos para la técnica de enseñanza 1, el valor $t_{.025}$ hubiera estado basado en sólo $n_1 - 1 = 5$ grados de libertad, el número de grados de libertad asociado con s_1. ■

EJEMPLO **13.4** Encuentre un intervalo de confianza de 95% para la diferencia en la calificación media para las técnicas de enseñanza 1 y 4, Ejemplo 13.2.

Solución El intervalo de confianza de 95% es

$$(\overline{Y}_{1\bullet} - \overline{Y}_{4\bullet}) \pm (2.093)\,(7.94)\sqrt{1/6 + 1/4} \quad \text{o} \quad -12.08 \pm 10.73.$$

En consecuencia, el intervalo de confianza de 95% para $(\mu_1 - \mu_4)$ es $(-22.81, -1.35)$. En el nivel de confianza de 95% concluimos que $\mu_4 > \mu_1$ por al menos 1.35 pero no más de 22.81. ■

Ejercicios

13.20 Consulte los Ejemplos 13.2 y 13.3.

a Use la parte de los datos de la Tabla 13.2 que se refiere sólo a la técnica de enseñanza 1 y el método de la Sección 8.8 para formar un intervalo de confianza de 95% para la calificación media de estudiantes instruidos usando la técnica 1.

b ¿Cómo se compara la longitud del intervalo de confianza de 95% hallada en el inciso a contra la longitud del intervalo de confianza de 95% obtenida en el Ejemplo 13.3?

c ¿Cuál es la principal razón para que el intervalo hallado en el inciso a sea más largo que el intervalo dado en el Ejemplo 13.3?

13.21 Consulte los Ejemplos 13.2 y 13.4.

a Use la parte de los datos de la Tabla 13.2 que se refiere sólo a las técnicas de enseñanza 1 y 4, así como el método de la Sección 8.8, para formar un intervalo de confianza de 95% para la diferencia en la calificación media para estudiantes instruidos usando las técnicas 1 y 4.

b ¿Cómo se compara la longitud del intervalo de confianza de 95% hallada en el inciso a con la longitud del intervalo de confianza de 95% obtenida en el Ejemplo 13.4?

c ¿Cuál es la principal razón para que el intervalo hallado en el inciso a sea más largo que el intervalo dado en el Ejemplo 13.4?

13.22 **a** Con base en sus respuestas a los Ejercicios 13.20 y 13.21, así como en los comentarios al final de esta sección, ¿cómo esperaría usted que se comparen los intervalos de confianza calculados usando los resultados de esta sección contra los intervalos relacionados que hacen uso de los datos de sólo una o dos de las muestras obtenidas en el diseño de un factor? ¿Por qué?

b Consulte el inciso a. ¿Es posible que un intervalo de confianza de 95% para la media de una población, basado sólo en la muestra tomada de dicha población, sea más corto que el intervalo de confianza de 95% para la misma media poblacional que se obtendría usando el procedimiento de esta sección? ¿Cómo?

13.23 Consulte el Ejercicio 13.7.

a Construya un intervalo de confianza de 95% para la cantidad media de aguas residuales contaminantes por galón para la planta A. Si el límite para la cantidad media de aguas residuales contaminantes es 1.5 libras/galón, ¿se concluiría que la planta A rebasa este límite? ¿Por qué?

b Dé un intervalo de confianza de 95% para la diferencia en la media de aguas residuales contaminantes por galón para las plantas A y D. ¿Este intervalo indica que la media de aguas residuales por galón difiere para estas dos plantas? ¿Por qué?

13.24 Consulte el Ejercicio 13.8. Construya un intervalo de confianza de 98% para la diferencia en la media de salarios iniciales para profesores auxiliares en instituciones públicas y privadas/independientes que otorgan doctorados.

13.25 Consulte el Ejercicio 13.11. Como se hizo notar en la descripción del experimento, los grupos con ooforectomía y premenopáusico de mujeres tenían aproximadamente la misma edad, pero las del grupo de ooforectomía sufrían de deficiencia de estrógeno. Forme un intervalo de confianza de 95% para la diferencia en densidades óseas medias para estos dos grupos de mujeres. ¿Concluiría que las densidades óseas medias para las mujeres con ooforectomía y premenopáusicas eran considerablemente diferentes? ¿Por qué?

13.26 Consulte el Ejercicio 13.9. Con μ_A y μ_B denote las resistencias medias de especímenes de concreto preparados con la mezcla A y la mezcla B, respectivamente.

a Determine un intervalo de confianza de 90% para μ_A.

b Determine un intervalo de confianza de 95% para $(\mu_A - \mu_B)$.

13.27 Consulte el Ejercicio 13.10. Con μ_A y μ_B, respectivamente, denote las calificaciones medias al final del semestre para las poblaciones de estudiantes hostiles en exceso que fueron tratados durante todo ese semestre por los métodos A y B, respectivamente. Encuentre un intervalo de confianza de 95% para

a $\mu_{A.}$

b μ_B.

c $(\mu_A - \mu_B)$.

13.28 Consulte el Ejercicio 13.12.

a Construya un intervalo de confianza de 95%, para el porcentaje medio de restos de carbono-14 en la pulpa de legumbre cuando se usa un nivel bajo de acetonitrilo.

b Dé un intervalo de confianza de 90%, para la diferencia en porcentajes medios de carbono-14 que quedan en la pulpa de legumbres, para niveles bajo y medio de acetonitrilo.

13.29 Consulte el Ejercicio 13.13.

a Proponga un intervalo de confianza de 95% para el tiempo medio de maniobra para dar vuelta a la izquierda para autobuses y camiones.

b Estime la diferencia en los tiempos medios de maniobra para autos pequeños y grandes, con un intervalo de confianza de 95%.

c El reporte del estudio hecho por Lu comprendía vehículos que pasaban por el crucero de Guadalupe Avenue y la Calle 38, en Austin, Texas. ¿Piensa usted que los resultados de los incisos a y b serían válidos para un crucero "desprotegido" en su ciudad? ¿Por qué sí o por qué no?

13.30 Se ha hecho la hipótesis de que tratamientos (después de moldeo) de un plástico empleado en lentes ópticos mejora las cualidades de resistencia al desgaste. Se han de probar cuatro tratamientos diferentes. Para determinar si existe alguna diferencia entre tratamientos se hicieron 28 moldes de una sola fórmula del plástico y se asignaron aleatoriamente 7 a cada uno de los tratamientos. La resistencia al desgaste se determinó al medir el aumento de "neblina" después de 200 ciclos de abrasión (una mejor resistencia al desgaste está indicada por menores aumentos). Los datos recolectados se indican en la siguiente tabla.

Tratamiento			
A	B	C	D
9.16	11.95	11.47	11.35
13.29	15.15	9.54	8.73
12.07	14.75	11.26	10.00
11.97	14.79	13.66	9.75
13.31	15.48	11.18	11.71
12.32	13.47	15.03	12.45
11.78	13.06	14.86	12.38

a ¿Hay evidencia de una diferencia en resistencia media al desgaste entre los cuatro tratamientos? Use $\alpha = .05$.

b Estime la diferencia media en el aumento de neblina entre los tratamientos B y C usando un intervalo de confianza de 99%.

c Encuentre un intervalo de confianza de 90% para la resistencia media al desgaste para lentes que reciben el tratamiento A.

13.31 Con la actual crisis energética, investigadores de las principales compañías petroleras están tratando de hallar fuentes alternativas de petróleo. Se sabe que algunos tipos de pizarra bituminosa contienen pequeñas cantidades de petróleo que es factible (mas no económico) extraer. Se han creado cuatro métodos para extraer petróleo de esta pizarra y el gobierno ha decidido que deben realizarse experimentos para determinar si los métodos difieren considerablemente en la cantidad promedio de petróleo que pueda extraerse de la pizarra con cada uno de ellos. Se sabe que el método 4 es el más costoso de instrumentar y que el método 1 es el menos costoso, de modo que las inferencias acerca de las diferencias en el rendimiento de estos dos métodos son de particular interés. Dieciséis trozos de pizarra (del mismo tamaño) se sometieron aleatoriamente a los cuatro métodos, con los resultados que se muestran en la siguiente tabla (las unidades son en litros por metro cúbico). Todas las inferencias han de hacerse con $\alpha = .05$.

Método 1	Método 2	Método 3	Método 4
3	2	5	5
2	2	2	2
1	4	5	4
2	4	1	5

a Suponiendo que las 16 unidades experimentales fueran lo más semejantes posible, realice el ANOVA apropiado para determinar si hay alguna diferencia importante entre las cantidades medias extraídas por los cuatro métodos. Use $\alpha = .05$.

b Construya un intervalo de confianza de 95% para la diferencia en las cantidades medias extraídas por los dos métodos de particular interés. Interprete el resultado.

13.32 Consulte el Ejercicio 13.14. Construya un intervalo de confianza de 95% para la cantidad media de residuo de DDT.

13.33 Consulte el Ejercicio 13.15. Compare el contenido medio de oxígeno disuelto a media corriente arriba de la planta contra el contenido medio adyacente a la planta (lugar 2 contra lugar 3). Use un intervalo de confianza de 95%.

13.34 Consulte el Ejercicio 13.15. Compare el contenido medio de oxígeno disuelto para los dos lugares aguas arriba de la planta, contra el contenido medio ligeramente aguas abajo de la planta, hallando un intervalo de confianza de 95% para $(1/2)\,(\mu_1 + \mu_2) - \mu_4$.

13.35 Consulte el Ejercicio 13.16. El promedio de aumento en el ritmo cardiaco para las diez personas de cada una de las categorías de edad fue

Edad	Tamaño muestral	Promedio de aumento en ritmo cardiaco
10–19	10	30.9
20–39	10	27.5
40–59	10	29.5
60–69	10	28.2

a Encuentre un intervalo de confianza de 90% para la diferencia en el aumento medio en la frecuencia cardiaca para los grupos de edades de 10–19 y 60–69.

b Encuentre un intervalo de confianza de 90% para el aumento medio en la frecuencia cardiaca para el grupo de edad de 20–39 años.

13.8 Modelo estadístico para el diseño de bloques aleatorizado

El método para construir un diseño de bloques aleatorizado se presentó en la Sección 12.4. Como ya indicamos en la Definición 12.6, el diseño de bloques aleatorizado se emplea para comparar k tratamientos usando b bloques. Los bloques se seleccionan de modo que, con optimismo, las unidades experimentales dentro de cada bloque sean homogéneas en esencia. Los tratamientos se asignan de manera aleatoria a las unidades experimentales de cada bloque, en forma tal que cada tratamiento aparece exactamente una vez en cada uno de los b bloques. Entonces, el número total de observaciones obtenidas en un diseño de bloques aleatorizado es $n = bk$. En el análisis de un diseño de bloques aleatorizado está implícita la presencia de dos variables cualitativas independientes, “bloques” y “tratamientos”. En esta sección presentamos un modelo estadístico formal para el diseño de bloques aleatorizado.

Modelo estadístico para un diseño de bloques aleatorizado

Para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, b$,

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

donde
Y_{ij} = observación del tratamiento i en el bloque j,
μ = la media general,
τ_i = efecto no aleatorio del tratamiento i, donde $\sum_{i=1}^{k} \tau_i = 0$,
β_j = el efecto no aleatorio del bloque j, donde $\sum_{j=1}^{b} \beta_j = 0$.
ε_{ij} = términos de error aleatorios tales que ε_{ij} son variables aleatorias independientes, distribuidas normalmente, con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$.

Observe que se supone que $\mu, \tau_1, \tau_2, \ldots, \tau_k$ y $\beta_1, \beta_2, \ldots, \beta_b$ son constantes que se suponen desconocidas. Este modelo difiere del de diseño completamente aleatorizado (un tipo específico de diseño de un factor) sólo en que contiene parámetros asociados con los diferentes bloques. Ya que se supone que los efectos de los bloques son fijos pero desconocidos, este modelo por lo general se conoce como modelo *de bloques* de *efectos fijos*. Un modelo de *efectos de bloque aleatorios*, otro modelo para el diseño de bloques aleatorizado en el que se supone que las β son variables aleatorias, se considera en los ejercicios complementarios. Nuestro desarrollo formal en el cuerpo de este texto está restringido al modelo de bloque de efectos fijos.

El modelo estadístico que acabamos de presentar resume de manera muy clara todas las suposiciones hechas en el análisis de datos en un diseño de bloques aleatorizado con efectos de bloque fijos. Consideremos la observación Y_{ij} hecha en el tratamiento i en el bloque j. Observe que las suposiciones del modelo implican que $E(Y_{ij}) = \mu + \tau_i + \beta_j$ y $V(Y_{ij}) = \sigma^2$ para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, b$. Consideremos las observaciones hechas en el tratamiento i y advirtamos que dos observaciones que reciben el tratamiento i tienen medias que difieren sólo por la diferencia de los efectos de bloque. Por ejemplo,

$$E(Y_{i1}) - E(Y_{i2}) = \mu + \tau_i + \beta_1 - (\mu + \tau_i + \beta_2) = \beta_1 - \beta_2.$$

Del mismo modo, dos observaciones que se tomen del mismo bloque tienen medias que difieren sólo por la diferencia de los efectos de tratamiento. Esto es, si $i \neq i'$,

$$E(Y_{ij}) - E(Y_{i'j}) = \mu + \tau_i + \beta_j - (\mu + \tau_{i'} + \beta_j) = \tau_i - \tau_{i'}.$$

Observaciones que se tomen en diferentes tratamientos y en diferentes bloques tienen medias que discrepan por la diferencia en los efectos de tratamiento más la diferencia en los efectos de bloque porque, si $i \neq i'$ y $j \neq j'$,

$$E(Y_{ij}) - E(Y_{i'j'}) = \mu + \tau_i + \beta_j - (\mu + \tau_{i'} + \beta_{j'}) = (\tau_i - \tau_{i'}) + (\beta_j - \beta_{j'}).$$

En la siguiente sección continuamos con un análisis de los datos obtenidos de un diseño de bloques aleatorizado.

Ejercicios

13.36 Exprese las suposiciones que sirven de base al ANOVA para un diseño de bloques aleatorizado con bloques de efectos fijos.

13.37 De acuerdo con el modelo para el diseño de bloques aleatorizado dado en esta sección, la respuesta esperada cuando el tratamiento i se aplica en el bloque j es $E(Y_{ij}) = \mu + \tau_i + \beta_j$, para $i = 1, 2, \dots, k$ y $j = 1, 2, \dots, b$.

a Use el modelo dado en esta sección para calcular el promedio de las $n = bk$ respuestas esperadas asociadas con todos los bloques y tratamientos.

b Dé una interpretación para el parámetro μ que aparece en el modelo para el diseño de bloques aleatorizado.

13.38 Con $\overline{Y}_{i\bullet}$ denotando el promedio de todas las respuestas al tratamiento i. Use el modelo para el diseño de bloques aleatorizado para deducir $E\left(\overline{Y}_{i\bullet}\right)$ y $V\left(\overline{Y}_{i\bullet}\right)$. ¿$\overline{Y}_{i\bullet}$ es un estimador insesgado para la respuesta media al tratamiento i? ¿Por qué sí o por qué no?

13.39 Consulte el Ejercicio 13.38 y considere $\overline{Y}_{i\bullet} - \overline{Y}_{i'}$ para $i \neq i'$.

a Demuestre que $E\left(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right) = \tau_i - \tau_{i'}$. Este resultado implica que $\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ es un estimador insesgado de la diferencia en los efectos de tratamiento i e i'.

b Deduzca $V\left(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right)$.

13.40 Consulte el modelo para el diseño de bloques aleatorizado y con $\overline{Y}_{\bullet j}$ denote el promedio de todas las respuestas del bloque j.

a Deduzca $E\left(\overline{Y}_{\bullet j}\right)$ y $V\left(\overline{Y}_{\bullet j}\right)$.

b Demuestre que $\overline{Y}_{\bullet j} - \overline{Y}_{\bullet j'}$ es un estimador insesgado para $\beta_j - \beta_{j'}$ la diferencia en los efectos de los bloques i y j'.

c Deduzca $V\left(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet j'}\right)$.

13.9 El análisis de varianza para el diseño de bloques aleatorizado

El ANOVA para un diseño de bloques aleatorizado tiene una forma muy semejante al de un diseño completamente aleatorizado (que es un caso especial del diseño de un factor). En el diseño de bloques aleatorizado, la suma total de cuadrados, SS total, se divide en tres partes: la suma de cuadrados por bloques, tratamientos y error.

Denote el total y promedio de todas las observaciones del bloque j como $Y_{\bullet j}$ y $\overline{Y}_{\bullet j}$, respectivamente. Del mismo modo, $Y_{i\bullet}$ y $\overline{Y}_{i\bullet}$ representan el total y el promedio de todas las observaciones que reciben el tratamiento i. De nuevo, los "puntos" en los subíndices indican cuál índice es el "sumado" para calcular los totales y el "promediado" para calcular los promedios. Entonces, para un diseño de bloques aleatorizado que contenga b bloques y k tratamientos tenemos las siguientes sumas de cuadrados:

$$\begin{aligned}
\text{SS total} &= \sum_{i=1}^{k}\sum_{j=1}^{b}(Y_{ij}-\overline{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{b}Y_{ij}^2 - \text{CM} \\
&= \text{SSB} + \text{SST} + \text{SSE}, \qquad \text{donde} \\
\text{SSB} &= k\sum_{j=1}^{b}(\overline{Y}_{\bullet j}-\overline{Y})^2 = \sum_{j=1}^{b}\frac{Y_{\bullet j}^2}{k} - \text{CM}, \\
\text{SST} &= b\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y})^2 = \sum_{i=1}^{k}\frac{Y_{i\bullet}^2}{b} - \text{CM}, \\
\text{SSE} &= \text{SS total} - \text{SSB} - \text{SST}.
\end{aligned}$$

En las fórmulas anteriores,

$$\overline{Y} = (\text{promedio de todas las } n = bk \text{ observaciones}) = \frac{1}{bk}\sum_{j=1}^{b}\sum_{i=1}^{k}Y_{ij},$$

y

$$\text{CM} = \frac{(\text{total de todas las observaciones})^2}{n} = \frac{1}{bk}\left(\sum_{j=1}^{b}\sum_{i=1}^{k}Y_{ij}\right)^2.$$

La tabla ANOVA para el diseño de bloques aleatorizado se presenta en la Tabla 13.5. Los grados de libertad asociados con cada suma de cuadrados se muestran en la segunda columna. Los cuadrados medios se calculan dividiendo la suma de cuadrados entre sus grados de libertad respectivos.

Para probar la hipótesis nula de que no hay diferencia en las medias de tratamiento usamos el estadístico F

Tabla 13.5 Tabla ANOVA para un diseño de bloques aleatorizado

Fuente	Grados de libertad	SS	MS
Bloques	$b-1$	SSB	$\dfrac{\text{SSB}}{b-1}$
Tratamientos	$k-1$	SST	$\dfrac{\text{SST}}{k-1}$
Error	$n-b-k+1$	SSE	MSE
Total	$n-1$	SS total	

$$F = \frac{\text{MST}}{\text{MSE}}$$

y rechazamos la hipótesis nula si $F > F_\alpha$, con base en $\nu_1 = (k-1)$ y $\nu_2 = (n-b-k+1)$ grados de libertad en numerador y denominador, respectivamente.

Como ya dijimos en la Sección 12.4, es posible usar bloqueo para controlar una fuente extraña de variación (la variación entre bloques). Además, con bloqueo tenemos la oportunidad de ver si existe evidencia para indicar una diferencia en la respuesta media para bloques. Dada la hipótesis nula de que no hay diferencia en la respuesta media para bloques (esto es, $\beta_j = 0$, para $j = 1, 2, \ldots, b$), el cuadrado medio para bloques (MSB) da un estimador insesgado para σ^2 con base en $(b-1)$ grados de libertad. En donde existan diferencias reales entre las medias de bloque, la MSB tenderá a estar influida en comparación con el MSE y

$$F = \frac{\text{MSB}}{\text{MSE}}$$

da un estadístico de prueba. Al igual que en la prueba para tratamientos, la región de rechazo para la prueba es

$$F > F_\alpha,$$

donde F tiene $\nu_1 = b-1$ y $\nu_2 = n-b-k+1$ grados de libertad en numerador y denominador, respectivamente.

EJEMPLO **13.5** Un experimento de respuesta a estímulo, que comprende tres tratamientos, se presentó en un diseño de bloques aleatorizado usando cuatro sujetos (personas). La respuesta fue el tiempo hasta la reacción, medido en segundos. Los datos se ven en bloques en la Figura 13.2. El número de tratamiento está circulado y se muestra arriba de cada observación. ¿Los datos presentan suficiente evidencia para indicar una diferencia en las respuestas medias para estímulos (tratamientos)? ¿Y en las personas? Use $\alpha = .05$ para cada prueba y dé los valores p asociados.

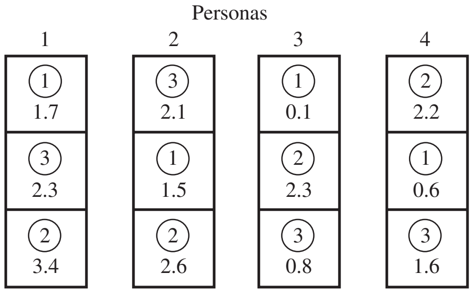


FIGURA **13.2** Diseño de bloques aleatorizado para el Ejemplo 13.5

Tabla 13.6 Tabla ANOVA para el Ejemplo 13.5

Fuente	Grados de libertad	SS	MS	F
Bloques	3	3.48	1.160	15.47
Tratamientos	2	5.48	2.740	36.53
Error	6	.45	.075	
Total	11	9.41		

Solución Los valores observados de las sumas de cuadrados para el ANOVA se muestran conjuntamente en la Tabla 13.6 y en forma individual como sigue:

$$\text{CM} = \frac{(\text{total})^2}{n} = \frac{(21.2)^2}{12} = 37.45,$$

$$\text{SS total} = \sum_{j=1}^{4}\sum_{i=1}^{3}(y_{ij} - \overline{y})^2 = \sum_{j=1}^{4}\sum_{i=1}^{3} y_{ij}^2 - \text{CM} = 46.86 - 37.45 = 9.41,$$

$$\text{SSB} = \sum_{j=1}^{4} \frac{Y_{\bullet j}^2}{3} - \text{CM} = 40.93 - 37.45 = 3.48,$$

$$\text{SST} = \sum_{i=1}^{3} \frac{Y_{i\bullet}^2}{4} - \text{CM} = 42.93 - 37.45 = 5.48,$$

$$\text{SSE} = \text{SS total} - \text{SSB} - \text{SST} = 9.41 - 3.48 - 5.48 = .45.$$

■

Usamos la relación entre MST y MSE para probar una hipótesis de que no hay diferencia en la respuesta media a los tratamientos. Así, el valor calculado de F es

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{2.74}{.075} = 36.53.$$

El valor crítico del estadístico $F(\alpha = .05)$ para $\nu_1 = 2$ y $\nu_2 = 6$ grados de libertad es $F_{.05} = 5.14$. Como el valor calculado de F excede al valor crítico, hay suficiente evidencia en el nivel $\alpha = .05$ para rechazar la hipótesis nula y concluir que existen diferencias reales entre las respuestas esperadas para los tres estímulos. El correspondiente valor $p = P(F > 36.53)$ que, con base en la Tabla 7, Apéndice 3, es tal que valor $p < .005$. La aplicación *F-Ratio Probabilities and Quantiles* da el valor $p = P(F > 36.53) = .00044$ exacto.

Es posible realizar una prueba similar para la hipótesis nula de que no existe diferencia en la respuesta media para personas. El rechazo de esta hipótesis implicaría que hay diferencias significativas entre personas y que es deseable el bloque. El valor calculado de F con base en $\nu_1 = 3$ y $\nu_2 = 6$ grados de libertad es

$$F = \frac{\text{MSB}}{\text{MSE}} = \frac{1.16}{.075} = 15.47.$$

Como este valor de F excede al correspondiente valor crítico tabulado, $F_{.05} = 4.76$, rechazamos la hipótesis nula y concluimos que existe una diferencia real en las respuestas medias entre los cuatro grupos de personas. La aplicación indica que el valor p asociado $= P(F > 15.47) = .00314$. Con base en la Tabla 7, Apéndice 3, habríamos concluido sólo que el valor $p < .005$. A pesar de todo, concluimos que el bloqueo por personas fue benéfico.

Ejercicios

13.41 En el Ejercicio 12.10 se realizó un análisis de observaciones pareadas para comparar las diferencias en tiempo medio de CPU para ejecutar programas de comparación en dos computadoras. Los datos se reproducen en la siguiente tabla.

	Programa de comparación					
Computadora	1	2	3	4	5	6
1	1.12	1.73	1.04	1.86	1.47	2.10
2	1.15	1.72	1.10	1.87	1.46	2.15

a Trate los seis programas como seis bloques y pruebe para una diferencia entre los tiempos medios de CPU para las dos computadoras usando un análisis de bloques aleatorizado. Use $\alpha = .05$. ¿Cómo se compara su decisión con la tomada en el Ejercicio 12.10(a)?

b Proporcione límites para el valor p asociado. ¿Cómo se compara su respuesta con su respuesta al Ejercicio 12.10(b)?

c **Ejercicio Applet** Use la aplicación *F-Ratio Probabilities and Quantiles* para hallar el valor p exacto.

d ¿Cómo se compara el valor del MSE con el valor para s_D^2 que usted empleó en su solución del Ejercicio 12.10?

13.42 La siguiente tabla presenta datos de producción que se relacionan con la resistencia a las manchas, para tres materiales (M_1, M_2 y M_3) tratados con cuatro productos químicos en un diseño de bloques aleatorizado. (Un valor bajo indica buena resistencia a las manchas.)

Producto químico	Material M_1	M_2	M_3	Total
A	5	9	7	21
B	3	8	4	15
C	8	13	9	30
D	4	6	8	18
Total	20	36	28	84
$\sum_i \sum_j y_{ij}^2 = 674$		$\frac{1}{12}\left(\sum_i \sum_j y_{ij}\right)^2 = 588$		

a ¿Hay evidencia de diferencias en la resistencia media entre los cuatro productos químicos? Dé límites para el valor p.

b ¿Qué se concluiría en el nivel de significancia $\alpha = .05$?

13.43 Consulte el Ejercicio 13.42. ¿Por qué se utilizó un diseño de bloques aleatorizado para comparar los productos químicos?

13.44 ¿El promedio de los costos de los seguros de automóviles difieren para diferentes compañías de seguros? Otras variables que afectan los costos de los seguros son la ubicación geográfica, edades de los conductores y el tipo de cobertura. Las siguientes son estimaciones (en dólares) del costo de pólizas de 6 meses para cobertura de responsabilidad básica para un solo hombre que ha tenido licencia de manejo durante 6–8 años, no ha cometido infracciones ni ha tenido accidentes y viaja en auto entre 12 600 y 15 000 millas por año.[6]

	Compañía de seguros				
Ubicación	21st Century	Allstate	AAA	Fireman's Fund	State Farm
Riverside	736	745	668	1065	1202
San Bernardino	836	725	618	869	1172
Hollywood	1492	1384	1214	1502	1682
Long Beach	996	884	802	1571	1272

a ¿Qué tipo de diseño se utilizó en la recolección de esta información?

b ¿Hay suficiente evidencia para indicar que el promedio de primas de seguro difiere de una compañía a otra?

c ¿Hay suficiente evidencia para indicar que las primas de seguro difieren de una ubicación a otra?

d **Ejercicio Applet** Use la aplicación *F-Ratio Probabilities and Quantiles* para hallar los valores p asociados con las pruebas en los incisos b y c.

13.45 Se realizó un experimento para determinar el efecto de tres métodos de preparación del suelo en el crecimiento de primer año de arbolitos de pino ayacahuite. Se seleccionaron cuatro lugares (terrenos de bosques del estado) y cada lugar se dividió en tres terrenos. Debido a que era probable que la fertilidad del suelo dentro de un terreno fuera más homogénea que entre lugares, se empleó un diseño de bloques aleatorizado, usando lotes como bloques. Los métodos de preparación del suelo fueron A (sin preparación), B (fertilización ligera) y C (quema). Cada preparación del suelo se aplicó a un lote dentro de cada lugar. En cada lote se plantó el mismo número de arbolitos y la observación registrada fue el promedio de crecimiento del primer año (en centímetros) de la semilla en cada lote. Estas observaciones se reproducen en la siguiente tabla.

6. *Fuente:* "2003 Auto Insurance," California Department of Insurance, http:cdinswww.insurance.ca.gov/pls/wu-survey-auto/apsw-get-prem$auto-mc.querylist, 23 de abril de 2004.

Preparación del suelo	Lugar			
	1	2	3	4
A	11	13	16	10
B	15	17	20	12
C	10	15	13	10

a Realice un ANOVA. ¿Los datos arrojan suficiente evidencia para indicar diferencias en el crecimiento medio para las tres preparaciones del suelo?

b ¿Hay evidencia para indicar diferencias en el crecimiento medio para los cuatro lugares?

13.46 A. E. Dudeck y C. H. Peacock informan sobre un experimento realizado para evaluar el rendimiento de varios céspedes de estación fría, para ver el efecto de siembra excesiva de *greens* de golf en invierno en el norte de la Florida. Una de las variables de interés fue la distancia que una pelota de golf rodaría en un *green* después de hacerla bajar por una rampa (que se usa para inducir una velocidad inicial constante a la pelota). Debido a que la distancia que la pelota rodaría estaba influida por la pendiente del *green* y la dirección en la que se podaba el césped, el experimento se realizó en un diseño de bloques aleatorizado. Los bloques se determinaron para que las pendientes de los tramos individuales fueran constantes dentro de bloques (se empleó un teodolito para asegurar la precisión) y todos los tramos fueron podados en la misma dirección y a la misma altura para eliminar los efectos de la poda. El césped base era "Tiftgreen Bermuda" en estado de casi hibernación. Se usaron el mismo método de siembra y rapidez de aplicación para todos los céspedes que están representados en la siguiente tabla de datos. Las medidas son distancias promedio (en metros) desde la base de la rampa hasta los puntos de parada de cinco pelotas que se hicieron bajar de la rampa y subir por la pendiente de cada tramo. Entre las variedades de césped usados en el estudio había A (ballico Pennfine), B (ballico Dasher), C (ballico real), D (supremo Marvelgreen) y E (ballico Barry). Las variedades de césped se plantaron dentro de bloques y dieron como resultado las medidas que se ilustran.[7]

	Variedad					
Bloque	A	B	C	D	E	Total
1	2.764	2.568	2.506	2.612	2.238	12.688
2	3.043	2.977	2.533	2.675	2.616	13.844
3	2.600	2.183	2.334	2.164	2.127	11.408
4	3.049	3.028	2.895	2.724	2.697	14.393
Total	11.456	10.756	10.268	10.175	9.678	52.333

a Realice el ANOVA apropiado para probar que hay suficiente evidencia para indicar que la distancia media rodada por la pelota difiere para las cinco variedades de césped. Indique límites para el nivel de significancia alcanzado. ¿Qué concluiría usted con un nivel de significancia de $\alpha = .01$?

b ¿Hay evidencia de una diferencia importante entre los bloques usados en el experimento? Pruebe usando $\alpha = .05$.

13.47 Consulte el Ejercicio 13.31. Suponga que ahora averiguamos que las 16 unidades experimentales se obtuvieron de la manera siguiente. Se tomó una muestra de cada uno de cuatro lugares, cada muestra individual se dividió en cuatro partes y luego cada método se aplicó a exactamente una parte de cada lugar (con la aleatorización apropiada). Los datos ahora se presentan de modo más correcto en la forma mostrada en la siguiente tabla. ¿Esta información sugiere un método más apropiado de análisis que

7. *Fuente:* A. E. Dudeck y C. H. Peacock, "Effects of Several Overseeded Ryegrasses on Turf Quality, Traffic Tolerance and Ball Roll," *Proceedings of the Fourth International Turfgrass Research Conference*, R. W. Sheard, ed., pp. 75–81. Ontario Agricultural College, University of Guelph, Guelph, Ontario, and the International Turfgrass Society, 1981.

el usado en el Ejercicio 13.31? Si es así, realice el nuevo análisis y conteste la pregunta del Ejercicio 13.31(a). ¿Es útil esta nueva información?

Lugar	Método 1	Método 2	Método 3	Método 4
I	3	2	5	5
II	2	2	2	2
III	1	4	5	4
IV	2	4	1	5

13.48 Suponga que a un diseño de bloques aleatorizado con b bloques y k tratamientos se le mide *dos veces* el tratamiento en cada bloque. Indique cómo haría usted los cálculos de un ANOVA.

13.49 Se realiza una evaluación de adherentes de difusión de componentes de aleación de circonio. El objetivo principal es determinar cuál de tres elementos: níquel, hierro o cobre, es el mejor agente de adherencia. Una serie de componentes de aleación de circonio se pegan usando cada uno de los posibles agentes de adherencia. Debido a una considerable variación en los componentes maquinados de diferentes lingotes, se emplea un diseño de bloques aleatorizado que bloquea los lingotes. Dos componentes de cada lingote se pegan usando cada uno de los tres agentes y se mide la presión (en unidades de 1000 libras por pulgada cuadrada) necesaria para separar los componentes pegados. Se obtienen los datos que se muestran en la tabla siguiente. ¿Hay evidencia de una diferencia en las presiones medias necesarias para separar los componentes entre los tres agentes adherentes? Use $\alpha = .05$.

	Agente adherente		
Lingote	Níquel	Hierro	Cobre
1	67.0	71.9	72.2
2	67.5	68.8	66.4
3	76.0	82.6	74.5
4	72.7	78.1	67.3
5	73.1	74.2	73.2
6	65.8	70.8	68.7
7	75.6	84.9	69.0

13.50 De vez en cuando, una sucursal de una empresa debe hacer envíos a otra sucursal en otro estado. Tres servicios de entrega de paquetería operan entre las dos ciudades donde están ubicadas las oficinas sucursales. Debido a que las estructuras de precios de los tres servicios de entrega son muy semejantes, la empresa desea comparar los tiempos de entrega. La empresa piensa hacer varios tipos de envíos diferentes a su sucursal. Para comparar los servicios de paquetería, la empresa remite cada envío por triplicado, uno con cada servicio de paquetería. Los resultados citados en la siguiente tabla son los tiempos de entrega en horas.

	Paquetería		
Envío	I	II	III
1	15.2	16.9	17.1
2	14.3	16.4	16.1
3	14.7	15.9	15.7
4	15.1	16.7	17.0
5	14.0	15.6	15.5

a ¿Hay evidencia de una diferencia en los tiempos medios de entrega entre los tres servicios de paquetería? Proporcione límites para el nivel de significancia alcanzado.

b ¿Por qué se realizó el experimento usando un diseño de bloques aleatorizado?

***13.51** Consulte el modelo de diseño de bloques aleatorizado presentado en la Sección 13.8.

a Deduzca $E(\text{MST})$.

b Deduzca $E(\text{MSB})$.

c Deduzca $E(\text{MSE})$.

Observe que estas cantidades aparecen en los estadísticos F usados para probar diferencias en la respuesta media entre los bloques y entre los tratamientos.

13.10 Estimación en el diseño de bloques aleatorizado

El intervalo de confianza para la diferencia entre un par de medias de tratamiento en un diseño de bloques aleatorizado es completamente análogo al asociado con el diseño completamente aleatorizado (un caso especial del diseño de un factor) de la Sección 13.7. Un intervalo de confianza de $100(1-\alpha)\%$ para $\tau_i - \tau_{i'}$ es

$$(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) \pm t_{\alpha/2} S \sqrt{\frac{2}{b}},$$

donde $n_i = n_{i'} = b$, el número de observaciones contenido en una media de tratamiento y $S = \sqrt{\text{MSE}}$. La diferencia entre los intervalos de confianza para los diseños completamente aleatorizado y de bloques aleatorizado es que el valor $t_{\alpha/2}$ está basado en $v = n - b - k + 1 = (b-1)(k-1)$ grados de libertad y que S, que aparece en la expresión anterior, se obtiene de la tabla ANOVA asociada con el diseño de bloques aleatorizado.

EJEMPLO **13.6** Construya un intervalo de confianza de 95% para la diferencia entre las respuestas medias a los tratamientos 1 y 2, Ejemplo 13.5.

Solución El intervalo de confianza para la diferencia en respuestas medias para un par de tratamientos es

$$(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) \pm t_{\alpha/2} S \sqrt{\frac{2}{b}},$$

donde el ejemplo $t_{.025}$ está basado en 6 grados de libertad. Para los tratamientos 1 y 2 tenemos

$$(.98 - 2.63) \pm (2.447)\,(.27)\sqrt{\frac{2}{4}}, \quad \text{o} \quad -1.65 \pm .47 = (-2.12, -1.18).$$

Entonces, con un nivel de confianza de 95%, concluimos que el tiempo medio de reacción para el estímulo 1 es entre 1.18 y 2.12 segundos menor que el tiempo medio de reacción para el estímulo 2. ■

Ejercicios

13.52 Consulte los Ejercicios 13.41 y 12.10. Encuentre un intervalo de confianza de 95% para la diferencia en tiempos medios de CPU necesarios para que las dos computadoras completen un trabajo. ¿Cómo se compara su respuesta con la obtenida en el Ejercicio 12.10(c)?

13.53 Consulte el Ejercicio 13.42. Construya un intervalo de confianza de 95% para la diferencia entre resistencias medias para los productos químicos A y B.

13.54 Consulte el Ejercicio 13.45. Construya un intervalo de confianza de 90% para las diferencias en crecimiento medio para los métodos A y B.

13.55 Consulte el Ejercicio 13.46. Construya un intervalo de confianza de 95% para la diferencia en la distancia media rodada cuando se usan céspedes Dasher y supremo Marvelgreen para siembra excesiva.

13.56 Consulte el Ejercicio 13.47. Construya un intervalo de confianza de 95% para la diferencia entre las cantidades medias de petróleo extraído por los métodos 1 y 4. Compare la respuesta con la obtenida en el Ejercicio 13.31(b).

13.57 Consulte el Ejercicio 13.49. Calcule la diferencia en las presiones medias para separar componentes que están pegados con níquel y hierro, usando un intervalo de confianza de 99%.

13.11 Selección del tamaño muestral

El método para seleccionar el tamaño muestral en el diseño de un factor (incluyendo el completamente aleatorizado) o el diseño de bloques aleatorizado es una extensión de los procedimientos de la Sección 8.7. Centraremos nuestra atención en el caso de tamaños muestrales iguales, $n_1 = n_2 = \cdots = n_k$, para tratamientos del diseño de un factor. El número de observaciones por tratamiento es igual al número de bloques b para el diseño de bloques aleatorizado. Entonces, el problema es determinar n_1 o b para estos dos diseños de manera que el experimento resultante contenga la cantidad de información que se desea.

La determinación de los tamaños muestrales sigue un procedimiento similar para ambos diseños; hacemos un resumen de un método general. Primero, el experimentador debe decidir sobre el parámetro (o parámetros) de interés principal. Por lo general esto comprende la comparación de un par de medias de tratamiento. En segundo término, el experimentador debe especificar un límite del error de estimación que pueda ser tolerado. Una vez que esto se haya determinado, lo siguiente es seleccionar n_i (el tamaño de la muestra de la población o tratamiento i) o bien, de manera correspondiente, b (el número de bloques para un diseño de bloques aleatorizado) que reducirá el semiancho del intervalo de confianza para el parámetro de modo que, en un nivel de confianza prescrito, sea menor o igual al límite especificado del error de estimación. Debe destacarse que la solución del tamaño muestral *siempre* será una aproximación porque σ es desconocida y una estimación para σ es desconocida hasta que se obtenga la muestra. La mejor estimación disponible para σ se usará para producir una solución aproximada. Ilustramos el procedimiento con un ejemplo.

EJEMPLO **13.7** Se elaborará un diseño completamente aleatorizado para comparar cinco técnicas de enseñanza en grupos de tamaños iguales. Se desea que la estimación de las diferencias en respuesta media en un examen de logros sea correcta hasta una variación de no más de 30 puntos de calificación del examen, con probabilidad igual a .95. Se espera que las calificaciones del examen para una técnica de enseñanza determinada posean un margen aproximadamente igual a 240. Encuentre el número aproximado de observaciones necesarias para cada muestra con el fin de obtener la información especificada.

Solución El intervalo de confianza para la diferencia entre un par de medias de tratamiento es

$$(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) \pm t_{\alpha/2} S \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}.$$

Por tanto, deseamos seleccionar n_i y $n_{i'}$ tal que

$$t_{\alpha/2} S \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}} \leq 30.$$

El valor de σ es desconocido y S es una variable aleatoria. No obstante, una solución aproximada para $n_i = n_{i'}$ se puede obtener si se conjetura que el valor observado de s será más o menos igual a un cuarto del margen. Así, $s \approx 240/4 = 60$. El valor de $t_{\alpha/2}$ estará basado en $(n_1 + n_2 + \cdots + n_5 - 5)$ grados de libertad e incluso para valores pares moderados de n_i, $t_{.025}$ será aproximadamente igual a 2. Entonces

$$t_{.025} s \sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}} \approx (2)(60)\sqrt{\frac{2}{n_i}} = 30,$$

o bien,

$$n_i = 32, \qquad i = 1, 2, \ldots, 5.$$

■

EJEMPLO **13.8** Se lleva a cabo un experimento para comparar los efectos tóxicos de tres sustancias químicas en la piel de ratas. Se esperaba que la resistencia a las sustancias químicas variara de manera importante de una rata a otra. Por tanto, las tres sustancias se habían de probar en cada rata para delinear las diferencias de una rata a otra.

La desviación estándar del error experimental era desconocida, pero experimentos anteriores que comprendían varias aplicaciones de una sustancia química similar en el mismo tipo de rata sugerían un margen de mediciones de respuesta igual a 5 unidades.

Encuentre un valor para b tal que el error en la estimación de la diferencia entre un par de medias de tratamiento es menor que 1 unidad, con probabilidad igual a .95.

Solución Un valor muy aproximado para s es un cuarto del margen, o sea $s \approx 1.25$. Entonces, deseamos seleccionar b para que

$$t_{.025} s \sqrt{\frac{1}{b} + \frac{1}{b}} = t_{.025} s \sqrt{\frac{2}{b}} \leq 1.$$

Como $t_{.025}$ dependerá de los grados de libertad asociados con s^2, que será $(n - b - k + 1)$, usaremos la aproximación $t_{.025} \approx 2$. Entonces,

$$(2)\,(1.25)\sqrt{\frac{2}{b}} = 1, \quad \text{o} \quad b \approx 13.$$

Aproximadamente trece ratas serán necesarias para obtener la información deseada. Como haremos tres observaciones ($k = 3$) por rata, nuestro experimento requerirá que se hagan un total de $n = bk = 13(3) = 39$ mediciones.

Los grados de libertad asociados con la estimación resultante s^2 serán $(n - b - k + 1) = 39 - 13 - 3 + 1 = 24$, con base en esta solución. Por tanto, el valor calculado de t parecería adecuado para esta solución aproximada. ■

Las soluciones de tamaño muestral para los Ejemplos 13.7 y 13.8 son muy aproximadas y tienen la intención de dar sólo una estimación poco precisa del tamaño muestral y de los costos consiguientes del experimento. Las longitudes reales de los intervalos de confianza resultantes dependerán de los datos que se observen en realidad. Estos intervalos pueden no tener las longitudes exactas especificadas por el experimentador pero tendrán el coeficiente de confianza necesario. Si los intervalos resultantes son todavía demasiado largos, el experimentador puede obtener información sobre σ cuando se recolecten los datos y puede volver a calcular una mejor aproximación al número de observaciones por tratamiento (n_i o b) a medida que avance el experimento.

Ejercicios

13.58 Consulte el Ejercicio 13.9.

a Aproximadamente, ¿cuántas muestras por mezcla de concreto deben prepararse para permitir el cálculo de la diferencia en las resistencias medias para un par preseleccionado de muestras hasta no más de .02 tonelada por pulgada cuadrada? Suponga el conocimiento de los datos dados en el Ejercicio 13.9.

b ¿Cuál es el número total de observaciones requerido en todo el experimento?

13.59 Consulte los Ejercicios 13.10 y 13.27(a). ¿Aproximadamente cuántas observaciones serían necesarias para que el error de estimación de μ_A no sea mayor que10 unidades? Use un coeficiente de confianza de 95%.

13.60 Consulte los Ejercicios 13.10 y 13.27(c).

a Suponiendo tamaños muestrales iguales para cada tratamiento, ¿aproximadamente cuántas observaciones del método A y del método B son necesarias para calcular $\mu_A - \mu_B$ con un error no mayor que 20 unidades? Use un coeficiente de confianza de 95%.

b ¿Cuál es el número total de observaciones requerido en todo el experimento?

13.61 Consulte el Ejercicio 13.45.

a ¿Cuántos lugares se necesitan usar para calcular la diferencia entre el crecimiento medio para cualesquiera dos preparaciones de suelo especificadas hasta no más de 1 unidad, con coeficiente de confianza de .95?

b ¿Cuál es el número total de observaciones necesarias en todo el experimento?

13.62 Consulte los Ejercicios 13.47 y 13.55. ¿Cuántos lugares deben usarse si se desea calcular $\mu_1 - \mu_4$ hasta no más de .5 unidad, con coeficiente de confianza .95?

13.12 Intervalos de confianza simultáneos para más de un parámetro

Los métodos de la Sección 13.7 se pueden usar para construir intervalos de confianza de $100(1 - \alpha)\%$ para una media de tratamiento individual o para la diferencia entre un par de medias de tratamiento en un diseño de un factor. Suponga que en el curso de un análisis deseamos construir varios de estos intervalos de confianza. El método de la Sección 13.10 se puede usar para comparar un par de medias de tratamiento en un diseño de bloques aleatorizado. Aun cuando es cierto que cada intervalo incluirá el parámetro estimado con probabilidad $(1-\alpha)$, ¿cuál es la probabilidad de que *todos* los intervalos contengan sus respectivos parámetros? El objetivo de esta sección es presentar un procedimiento para formar conjuntos de intervalos de confianza para que el coeficiente de confianza *simultáneo* sea no menor que $(1-\alpha)$ para cualquier valor especificado de α.

Suponga que deseamos hallar intervalos de confianza $I_1, I_2, \dots, I_m$ para parámetros $\theta_1, \theta_2, \dots, \theta_m$ para que

$$P(\theta_j \in I_j \text{ para toda } j = 1, 2, \dots, m) \geq 1-\alpha.$$

Esta meta se puede alcanzar si se usa una desigualdad de probabilidad simple, conocida como *desigualdad de Bonferroni* (recuerde el Ejercicio 2.104). Para cualesquiera eventos $A_1, A_2, \dots, A_m$, tenemos

$$\overline{A_1 \cap A_2 \cap \cdots \cap A_m} = \overline{A}_1 \cup \overline{A}_2 \cup \cdots \cup \overline{A}_m.$$

Por tanto,

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) = 1 - P(\overline{A}_1 \cup \overline{A}_2 \cup \cdots \cup \overline{A}_m).$$

También, de la ley aditiva de probabilidad sabemos que

$$P(\overline{A}_1 \cup \overline{A}_2 \cup \cdots \cup \overline{A}_m) \leq \sum_{j=1}^{m} P(\overline{A}_j).$$

En consecuencia, obtenemos la *desigualdad de Bonferroni*

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) \geq 1 - \sum_{j=1}^{m} P(\overline{A}_j).$$

Suponga que $P(\theta_j \in I_j) = 1 - \alpha_j$ y sea A_j el evento $\{\theta_j \in I_j\}$. Entonces,

$$P(\theta_1 \in I_1, \dots, \theta_m \in I_m) \geq 1 - \sum_{j=1}^{m} P(\theta_j \notin I_j) = 1 - \sum_{j=1}^{m} \alpha_j.$$

Si todas las α_j, para $j = 1, 2, \dots, m$, se seleccionan iguales a α, podemos ver que el coeficiente de confianza simultáneo de los intervalos I_j, para $j = 1, 2, \dots, m$, podría ser de sólo $(1 - m\alpha)$, que es menor que $(1-\alpha)$ si $m > 1$. Un coeficiente simultáneo de confianza de al menos $(1-\alpha)$ puede asegurarse si se seleccionan los intervalos de confianza I_j, para $j = 1, 2, \dots, m$, para que $\sum_{j=1}^{m} \alpha_j = \alpha$. Una forma de lograr este objetivo es que cada intervalo se construya para que tenga coeficiente de confianza $1 - (\alpha/m)$. Aplicamos esta técnica en el siguiente ejemplo.

EJEMPLO 13.9 Para los cuatro tratamientos dados en el Ejemplo 13.2 construya intervalos de confianza para todas las comparaciones de la forma $\mu_i - \mu_{i'}$, con coeficiente simultáneo de confianza no menor que .95.

Solución El intervalo de confianza apropiado de $100(1-\alpha)\%$ para una sola comparación (por ejemplo $(\mu_1-\mu_2)$ es

$$(\overline{Y}_{1\bullet}-\overline{Y}_{2\bullet}) \pm t_{\alpha/2}S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

Como hay seis de estas diferencias por considerar, cada intervalo debe tener coeficiente de confianza $1-(\alpha/6)$. Entonces, el correspondiente valor t es $t_{\alpha/2(6)} = t_{\alpha/12}$. Como deseamos que el coeficiente de confianza simultáneo sea al menos .95, el valor t apropiado es $t_{.05/12} = t_{.00417}$. Usando la Tabla 5, Apéndice 3, el valor de tabla disponible más cercano es $t_{.005}$, de modo que usaremos éste para calcular el resultado deseado. El MSE para los datos del Ejemplo 13.2 está basado en 19 grados de libertad, de modo que el valor en la tabla es $t_{.005} = 2.861$.

Como $s = \sqrt{\text{MSE}} = \sqrt{63} = 7.937$, el intervalo para $\mu_1-\mu_2$ entre los seis con coeficiente de confianza simultáneo de al menos .95 es

$$\mu_1-\mu_2\text{: } (75.67-78.43) \pm 2.861(7.937)\sqrt{\frac{1}{6}+\frac{1}{7}} \quad \text{o} \quad -2.76 \pm 12.63.$$

De manera análoga, el conjunto completo de seis intervalos realizados es

$$\begin{aligned}
\mu_1-\mu_2&: \quad -2.76 \pm 12.63\\
\mu_1-\mu_3&: \quad 4.84 \pm 13.11\\
\mu_1-\mu_4&: \quad -12.08 \pm 14.66\\
\mu_2-\mu_3&: \quad 7.60 \pm 12.63\\
\mu_2-\mu_4&: \quad -9.32 \pm 14.23\\
\mu_3-\mu_4&: \quad -16.92 \pm 14.66.
\end{aligned}$$

No podemos lograr nuestro objetivo de obtener un conjunto de seis intervalos de confianza con coeficiente simultáneo de confianza de al menos .95 porque las tablas t del texto son demasiado limitadas. Desde luego, existen tablas más extensas de las distribuciones t. Como cada uno de nuestros seis intervalos tiene coeficiente de confianza .99, podemos decir que los seis intervalos anteriores tienen un coeficiente de confianza simultáneo de al menos .94. La aplicación *Student's t Probabilities and Quantiles*, aplicado con 19 grados de libertad, da $t_{.00417} = 2.9435$. Se pueden obtener intervalos con coeficiente de confianza simultánea de .9499 al sustituir $t_{.00417} = 2.9435$ en lugar de 2.861 en los cálculos previos. ■

Destacamos que la técnica presentada en esta sección garantiza probabilidades simultáneas de al menos $1-\alpha$. La probabilidad simultánea real puede ser mucho mayor que el valor nominal $1-\alpha$. Otros métodos para construir intervalos de confianza simultáneos se pueden hallar en los libros citados en la bibliografía que está al final de este capítulo.

Ejercicios

13.63 Consulte el Ejemplo 13.9. Los seis intervalos de confianza para $\mu_i-\mu_{i'}$ se obtuvieron mediante el uso de un valor aproximado (debido a la limitación de la información en la Tabla 5, Apéndice 3) para $t_{.00417}$. ¿Por qué difieren en longitud algunos intervalos?

13.64 Consulte el Ejercicio 13.63 y el Ejemplo 13.9.

a Use el valor exacto para $t_{.00417}$ dado en el Ejemplo 13.9 para dar un intervalo de 99.166% para $\mu_1 - \mu_2$. Este intervalo es uno de los seis intervalos simultáneos para $\mu_i - \mu_{i'}$ con coeficiente de confianza simultáneo no menor que .94996 ≈ .95.

b ¿Cuál es la razón de las longitudes de los intervalos para $\mu_1 - \mu_2$ obtenidas en el Ejemplo 13.9 y el inciso a?

c ¿Cómo se compara la razón que usted obtuvo en el inciso b con la razón $t_{.005}/t_{.00417}$?

d Con base en los incisos b, c y el intervalo para $\mu_1 - \mu_3$ dado en el Ejemplo 13.9 proporcione un intervalo de 99.166% para $\mu_1 - \mu_3$. Al igual que antes, éste es uno de los seis intervalos simultáneos para comparar μ_i y $\mu_{i'}$ con coeficiente de confianza simultáneo no menor que .94996 ≈ .95.

13.65 Consulte el Ejercicio 13.13. Construya intervalos de confianza para todas las diferencias posibles entre los tiempos medios de maniobra para las tres clases de vehículos, de modo que el coeficiente simultáneo de confianza sea al menos .95. Interprete los resultados.

13.66 Consulte el Ejercicio 13.12. Después de ver los datos, un lector del informe de Wheeler y colegas notó que la diferencia máxima entre las medias muestrales ocurre cuando se comparan concentraciones altas y bajas de acetonitrilo. Si se desea un intervalo de confianza para la diferencia en las medias poblacionales correspondientes, ¿cómo sugeriría usted construir este intervalo?

13.67 Consulte el Ejercicio 13.45. Construya intervalos de confianza para todas las diferencias posibles entre las medias de tratamiento (preparación del suelo), para que el coeficiente simultáneo de confianza sea al menos .90.

13.68 Consulte los Ejercicios 13.31 y 13.47. Como el método 4 es el más costoso, se desea compararlo contra los otros tres. Construya intervalos de confianza para las diferencias $\mu_1 - \mu_4$, $\mu_2 - \mu_4$ y $\mu_3 - \mu_4$ para que el coeficiente simultáneo de confianza sea al menos .95.

13.13 Análisis de varianza usando modelos lineales

Los métodos para analizar los modelos lineales presentados en el Capítulo 11 se pueden adaptar para su uso en el ANOVA. Ilustramos el método al formular un modelo lineal para datos obtenidos por medio de un diseño completamente aleatorizado que comprende $k = 2$ tratamientos.

Sea Y_{ij} la variable aleatoria en la j-ésima observación del tratamiento i, para $i = 1, 2$. Definamos una *variable x ficticia* o *indicadora*, de la siguiente manera:

$$x = \begin{cases} 1, & \text{si la observación es de la población 1.} \\ 0, & \text{de otro modo.} \end{cases}$$

Aunque dichas variables mudas se pueden definir en numerosas formas, esta definición es consistente con la codificación empleada en el SAS y otros programas de computadora para análisis estadístico. Observe que con esta codificación x es 1 si la observación se toma de la población 1 y x es 0 si la observación se toma de la población 2. Si usamos x como una variable independiente en un modelo lineal, podemos modelar Y_{ij} como

$$Y_{ij} = \beta_0 + \beta_1 x + \varepsilon_{ij},$$

donde ε_{ij} es un error aleatorio distribuido normalmente con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma^2$. En este modelo,

$$\mu_1 = E(Y_{1j}) = \beta_0 + \beta_1(1) = \beta_0 + \beta_1,$$

y

$$\mu_2 = E(Y_{2j}) = \beta_0 + \beta_1(0) = \beta_0.$$

Entonces, se deduce que $\beta_1 = \mu_1 - \mu_2$ y una prueba de la hipótesis $\mu_1 - \mu_2 = 0$ es equivalente a la prueba de que $\beta_1 = 0$. Nuestra intuición sugeriría que $\hat{\beta}_0 = \overline{Y}_{2\bullet}$ y $\hat{\beta}_1 = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$ son buenos estimadores de β_0 y β_1; de hecho, se puede demostrar (prueba omitida) que éstos son los estimadores de mínimos cuadrados obtenidos al ajustar el modelo lineal anterior. Ilustramos el uso de esta técnica volviendo a analizar los datos presentados en el Ejemplo 13.1.

EJEMPLO **13.10** Ajuste un modelo lineal apropiado a los datos del Ejemplo 13.1 y haga una prueba para ver si hay una diferencia importante entre μ_1 y μ_2.

Solución El modelo, como ya indicamos antes, está dado por

$$Y_{ij} = \beta_0 + \beta_1 x + \varepsilon_{ij},$$

donde

$$x = \begin{cases} 1, & \text{si la observación es de la población 1,} \\ 0, & \text{de otro modo.} \end{cases}$$

Las matrices empleadas para los estimadores de mínimos cuadrados son entonces

$$\mathbf{Y} = \begin{bmatrix} 6.1 \\ 7.1 \\ 7.8 \\ 6.9 \\ 7.6 \\ 8.2 \\ 9.1 \\ 8.2 \\ 8.6 \\ 6.9 \\ 7.5 \\ 7.9 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix},$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 6 \\ 6 & 6 \end{bmatrix}, \qquad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/6 & -1/6 \\ -1/6 & 1/3 \end{bmatrix}.$$

Los estimadores de mínimos cuadrados están dados por

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1/6 & -1/6 \\ -1/6 & 1/3 \end{bmatrix} \begin{bmatrix} 91.9 \\ 43.7 \end{bmatrix} = \begin{bmatrix} 8.033 \\ -.75 \end{bmatrix}.$$

Observe que $\hat{\beta}_0 = 8.033 = \overline{Y}_{2\bullet}$ y $\hat{\beta}_1 = -.75 = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$.

Además,

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 5.8617$$

es igual que la SSE calculada en el Ejemplo 13.1. Por tanto, $s^2 = \text{SSE}/(n-2) = .58617$ y $s = \sqrt{.58617} = .7656$.

Para probar $H_0 : \beta_1 = 0$, construimos el estadístico t (véase la Sección 11.12):

$$t = \frac{\hat{\beta}_1 - 0}{s\sqrt{c_{11}}} = \frac{-.75}{.7656\sqrt{1/3}} = -1.697.$$

Como estamos interesados en una prueba de dos colas, el valor p asociado es $2P(t < -1.697) = 2P(t > 1.697)$, donde t está basada en 10 grados de libertad. Entonces, usando la Tabla 5, Apéndice 3, obtenemos $.05 < P(t > 1.697) < .10$ y $.10 <$ valor $p < .20$. Por tanto, para cualquier valor α menor que 0.1, no podemos rechazar H_0. Esto es, hay suficiente evidencia para indicar que μ_1 y μ_2 difieren.

Esta prueba t es equivalente a la prueba F del Ejemplo 13.1. De hecho, el cuadrado del valor t observado es el valor F observado en el Ejemplo 13.1. ■

Ilustramos el método del modelo lineal para un análisis más complicado de problema de varianza al considerar un diseño de bloques aleatorizado.

EJEMPLO **13.11** Se realizó un experimento para comparar los efectos de cuatro sustancias químicas A, B, C y D en la resistencia al agua en textiles. Se usaron tres rollos diferentes de material I, II y III, con cada uno de los tratamientos químicos aplicado a una pieza de material cortado de cada uno de los rollos. Los datos se dan en la Tabla 13.7. Escriba un modelo lineal para este experimento y pruebe la hipótesis de que no hay diferencias entre las resistencias medias al agua para las cuatro sustancias químicas. Use $\alpha = .05$.

Solución Al formular el modelo, definimos β_0 como la respuesta media al tratamiento D en el material del rollo III y luego introducimos una variable indicadora distinta para cada tratamiento y para cada rollo de material (bloque). El modelo es

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon,$$

donde

$$x_1 = \begin{cases} 1, & \text{si se usa material del rollo I,} \\ 0, & \text{de otro modo,} \end{cases}$$

$$x_2 = \begin{cases} 1, & \text{si se usa material del rollo II,} \\ 0, & \text{de otro modo,} \end{cases}$$

$$x_3 = \begin{cases} 1, & \text{si se usa el tratamiento A,} \\ 0, & \text{de otro modo,} \end{cases}$$

Tabla 13.7 Datos para el Ejemplo 13.11

	Tratamientos			
Rollo de material	A	B	C	D
I	10.1	11.4	9.9	12.1
II	12.2	12.9	12.3	13.4
III	11.9	12.7	11.4	12.9

$$x_4 = \begin{cases} 1, & \text{si se usa el tratamiento B,} \\ 0, & \text{de otro modo,} \end{cases}$$
$$x_5 = \begin{cases} 1, & \text{si se usa el tratamiento C,} \\ 0, & \text{de otro modo.} \end{cases}$$

Buscamos probar la hipótesis de que no hay diferencias entre las medias de tratamiento, que es equivalente a $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. Así, debemos ajustar un modelo completo y uno reducido. (Vea la Sección 11.14.)

Para el modelo completo, tenemos

$$\mathbf{Y} = \begin{bmatrix} 10.1 \\ 12.2 \\ 11.9 \\ 11.4 \\ 12.9 \\ 12.7 \\ 9.9 \\ 12.3 \\ 11.4 \\ 12.1 \\ 13.4 \\ 12.9 \end{bmatrix} \quad \text{y} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Un poco de álgebra de matrices da como resultado, para este modelo completo,

$$\text{SSE}_C = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 1721.760 - 1721.225 = .535.$$

El modelo reducido relevante es

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

y la matriz **X** correspondiente está formada por sólo las primeras tres columnas de la matriz **X** dada para el modelo completo. Entonces obtenemos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 12.225 \\ -1.350 \\ .475 \end{bmatrix}$$

y

$$\text{SSE}_R = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} = 1721.760 - 1716.025 = 5.735.$$

Se deduce que la relación F apropiada para comparar estos modelos completo y reducido es

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{\text{SSE}_C/(n-[k+1])} = \frac{(5.735 - .535)/(5-2)}{(.535)/(12-6)} = \frac{1.733}{.0892} = 19.4.$$

La F tabulada para $\alpha = .05$, $\nu_1 = 3$, y $\nu_2 = 6$ es 4.76. En consecuencia, si escogemos $\alpha = .05$, rechazamos la hipótesis nula y concluimos que los datos presentan suficiente evidencia para indicar que existen diferencias entre las medias de tratamiento. El valor p asociado está dado por $P(F > 19.4)$. La Tabla 7, Apéndice 3, establece que el valor $p < .005$. La aplicación *F-Ratio Probabilities and Quantiles*, aplicada con 3 grados de libertad en el numerador y 6 grados de libertad en el denominador dan valor $p = P(F > 19.4) = .00172$. La prueba F empleada en este ejemplo es equivalente a la que hubiera sido producida por los métodos explicados en la Sección 13.9. ■

A pesar de ser una técnica muy útil, el método del modelo lineal para el cálculo del ANOVA se usa por lo general sólo cuando los cálculos se hacen en computadora. Las fórmulas de cálculo indicadas anteriormente en el capítulo son más convenientes cuando se hacen cálculos manualmente. Observe que si hay k tratamientos comprendidos en un estudio, el método de las "variables ficticias" requiere que definamos $k - 1$ variables ficticias si deseamos usar el método del modelo lineal para analizar los datos.

Ejercicios

13.69 Consulte el Ejemplo 13.11. En el Ejercicio 13.37 usted interpretó los parámetros del modelo para un diseño de bloques aleatorizado en términos de la respuesta media para cada tratamiento en cada bloque. En términos del modelo con variables ficticias dado en el Ejemplo 13.11, β_0 es la respuesta media al tratamiento D para el rollo de material (bloque) III.

a En términos de los valores β, ¿cuál es la respuesta media al tratamiento A del bloque III?

b Con base en su respuesta al inciso a, ¿cuál es una interpretación del parámetro β_3?

13.70 Consulte el Ejercicio 13.10.

a Conteste la pregunta planteada en el Ejercicio 13.10 ajustando los modelos lineales completo y reducido. Pruebe usando $\alpha = .05$.

b Use los cálculos para el modelo completo del inciso a para probar la hipótesis de que no hay diferencia entre las medias para los métodos A y C. Pruebe usando $\alpha = .05$.

c Indique los niveles de significancia alcanzados para las pruebas realizadas en los incisos a y b.

13.71 Consulte el Ejercicio 13.42. Conteste el inciso a al ajustar los modelos completo y reducido.

13.72 Consulte el Ejercicio 13.45. Conteste el inciso b al construir una prueba F usando los modelos lineales completo y reducido.

13.14 Resumen

El diseño de un factor (incluyendo el diseño completamente aleatorizado) y el diseño de bloques aleatorizado son ejemplos de experimentos que comprenden una y dos variables cualitativas independientes, respectivamente. El ANOVA divide las sumas totales de cuadrados, SS total, en partes asociadas con cada variable independiente y error experimental. Los cuadrados medios asociados con cada variable independiente pueden compararse con el MSE para ver si el valor de los cuadrados medios son suficientemente grandes como para implicar que la variable independiente tiene algún efecto sobre la respuesta. Los intervalos de confianza para la respuesta media a un tratamiento individual o la diferencia en las respuestas medias para dos tratamientos preseleccionados son modificaciones sencillas de intervalos presentadas en capítulos anteriores. La desigualdad de Bonferroni se utilizó para construir un conjunto de intervalos de confianza con *coeficiente de confianza simultáneo* de al menos $1 - \alpha$. Finalmente, introdujimos el método de variable ficticia que permite el uso de metodología de modelos lineales para realizar un análisis de varianza.

En este capítulo hemos presentado una introducción muy breve al análisis de varianza y su tema asociado, el diseño de experimentos. Pueden diseñarse experimentos para investigar el efecto de numerosas variables cuantitativas y cualitativas en una respuesta. Éstas pueden ser variables de gran interés para el experimentador, así como variables de ruido como son los bloques, que pueden contribuir a una variación no deseada que tratamos de separar del error experimental. Cuando se diseñan de modo apropiado, estos experimentos aportan datos que pueden ser analizados usando un método ANOVA. Un análisis más amplio de los conceptos básicos de diseño experimental y del análisis de experimentos está en la bibliografía.

Bibliografía y lecturas adicionales

Box, G. E. P., W. G. Hunter, and J. S. Hunter. 2005. *Statistics for Experimenters,* 2d ed. New York: Wiley Interscience.

Cochran, W. G., and G. Cox. 1992. *Experimental Designs*, 2d ed. New York: Wiley.

Graybill, F. 2000. *Theory and Application of the Linear Model*. Belmont Calif.: Duxbury.

Hicks, C. R., and K. V. Turner. 1999. *Fundamental Concepts in the Design of Experiments*, 5th ed. New York: Oxford University Press.

Hocking, R. R. 2003. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 5th ed. New York: Wiley Interscience.

Montgomery, D. C. 2006. *Design and Analysis of Experiments*, 6th ed. New York: Wiley.

Scheaffer, R. L., W. Mendenhall, and L. Ott. 2006. *Elementary Survey Sampling*, 6th ed. Belmont Calif.: Duxbury.

Scheffé, H. 2005. *The Analysis of Variance.* New York: Wiley Interscience.

Ejercicios complementarios

13.73 Suponga que $n = bk$ unidades experimentales se encuentran disponibles para ser utilizadas en un experimento empleado para comparar k tratamientos. Si se pueden formar bloques en una forma significativa, ¿cómo deben ser identificadas las unidades experimentales de cada bloque?

13.74 Consulte el Ejercicio 13.73.

a Si se emplea un diseño completamente aleatorizado, ¿cómo seleccionaría usted las unidades experimentales que están asignadas a los tratamientos diferentes?

b Si se emplea un diseño de bloques aleatorizado, ¿cómo seleccionaría usted las unidades experimentales que están asignadas a cada uno de los k tratamientos?

13.75 Tres agentes para limpiar la piel se usaron en tres personas. Para cada persona, tres áreas de piel se expusieron a un contaminante y después se limpiaron usando uno de los tres agentes de limpieza. Después de 8 horas se midió el contaminante residual, con los siguientes resultados:

$$\text{SST} = 1.18, \qquad \text{SSB} = .78, \qquad \text{SSE} = 2.24.$$

a ¿Cuáles son las unidades experimentales y cuáles son los bloques en este experimento?

b Pruebe la hipótesis de que no hay diferencias entre las medias de tratamiento, usando $\alpha = .05$.

13.76 Consulte el Ejercicio 13.9. Suponga que la arena utilizada en las mezclas para las muestras 1–4 provino de la fosa A, la de las muestras 5–8 de la fosa B y la de las muestras 9–12 de la fosa C. Analice los datos suponiendo que las necesidades para un bloque aleatorizado son satisfechas con tres bloques consistentes, respectivamente, en las muestras 1, 2, 3 y 4; en las muestras 5, 6, 7 y 8; y en las muestras 9, 10, 11 y 12.

a En el nivel de significancia de 5%, ¿hay evidencia de diferencias en resistencia del concreto debido a la arena usada?

b ¿Hay evidencia, en el nivel de significancia de 5%, de diferencias en las resistencias promedio entre los cuatro tipos de concreto usado?

c ¿La conclusión del inciso b contradice la conclusión obtenida en el Ejercicio 13.9?

13.77 Consulte el Ejercicio 13.76. Sean μ_A y μ_B, respectivamente, las resistencias medias de las muestras de concreto preparadas de la mezcla A y la mezcla B.

a Encuentre un intervalo de confianza de 95% para $(\mu_A - \mu_B)$.

b El intervalo encontrado en el inciso a ¿es el mismo intervalo hallado en el Ejercicio 13.26(b)? ¿Por qué sí o por qué no?

13.78 Se inició un estudio para investigar el efecto de dos medicamentos, administrados simultáneamente, para reducir la presión sanguínea en personas. Se decidió usar tres niveles de cada medicamento e incluir las nueve combinaciones en el experimento. Nueve pacientes con presión alta fueron seleccionados para el experimento y aleatoriamente se asignó uno a cada una de las nueve combinaciones de medicamento. La respuesta observada fue una reducción en la presión sanguínea en un intervalo fijo de tiempo.

a ¿Es este un diseño de bloques aleatorizado?

b Suponga que dos pacientes se asignaron aleatoriamente a cada una de las nueve combinaciones de medicamento. ¿Qué tipo de diseño experimental es este?

13.79 Consulte el Ejercicio 13.78. Suponga que se empleará un diseño balanceado y completamente aleatorizado y que experimentos anteriores sugieren que $\sigma = 20$.

a ¿Cuántas repeticiones serían necesarias para calcular cualquier media de tratamiento (combinación de medicamento) correcta, con variación de no más de ± 10 con probabilidad .95?

b ¿De cuántos grados de libertad se dispondrá para calcular σ^2 cuando se use el número de repeticiones determinado en el inciso a?

c Determine la mitad de la amplitud aproximada de un intervalo de confianza de 95% para la diferencia en las respuestas medias para dos tratamientos cuando se use el número de repeticiones determinado en el inciso a.

13.80 Un distribuidor tiene en existencia tres autos (modelos A, B y C) de la misma marca pero diferentes modelos. Con el deseo de comparar el rendimiento en millas obtenido de estos diferentes modelos, un cliente hizo arreglos para probar cada uno de los autos con cada una de tres marcas de gasolina (marcas X, Y y Z). En cada intento, se vertió un galón de gasolina en el tanque vacío y el auto se hizo funcionar sin detenerse hasta agotar la gasolina. La tabla siguiente muestra el número de millas recorridas en cada uno de los nueve intentos.

	Distancia (millas)		
Marca de gasolina	Modelo A	Modelo B	Modelo C
X	22.4	17.0	19.2
Y	20.8	19.4	20.2
Z	21.5	18.7	21.2

a ¿El cliente debe concluir que los diferentes modelos de autos dan distinto rendimiento medio de gasolina? Pruebe con un nivel $\alpha = .05$?

b ¿Los datos indican que la marca de gasolina afecta el rendimiento de ésta?

13.81 Consulte el Ejercicio 13.80. Suponga que el rendimiento de la gasolina no está relacionado con su marca. Realice un análisis apropiado de los datos para un diseño completamente aleatorizado con tres tratamientos.

a ¿Debe el cliente concluir que los tres autos difieren en rendimiento de gasolina? Pruebe en el nivel $\alpha = .05$.

b Comparando su respuesta para el Ejercicio 13.80(a) con su respuesta para el inciso a de este ejercicio, ¿puede sugerir una razón por la cual un diseño en bloques puede no ser inteligente en ciertos casos?

c ¿Por qué podría ser un *error* analizar los datos en la forma sugerida en el inciso a?

13.82 Con la esperanza de atraer más viajeros, una compañía de transportes urbanos planea tener servicio de autobuses express de una terminal suburbana al centro financiero de la ciudad. Estos autobuses deben ahorrar tiempo de viaje. La ciudad decide realizar un estudio del efecto de cuatro planes diferentes (por ejemplo un carril especial y semáforos sincronizados) en el tiempo de viaje de los autobuses. Los tiempos de viaje (en minutos) se miden para varios días hábiles durante un viaje en horas pico cuando cada uno de los planes se está aplicando. Los resultados aparecen en la tabla siguiente.

Plan			
1	2	3	4
27	25	34	30
25	28	29	33
29	30	32	31
26	27	31	
	24	36	

a ¿Qué tipo de diseño experimental se utilizó?

b ¿Hay evidencia de una diferencia en los tiempos medios de viaje para los cuatro planes? Use $\alpha = 0.01$.

c Forme un intervalo de confianza de 95% para la diferencia entre el plan 1 (carril express) y el plan 3 (un control: no hay arreglos especiales de viaje).

13.83 Se realizó un estudio para comparar el efecto de tres niveles de digitalina en el nivel de calcio en el músculo cardiaco de perros. Se omite una descripción del procedimiento experimental real, pero es suficiente tomar nota de que el nivel general de ingesta de calcio varía de un animal a otro, de modo que la comparación de los niveles de digitalina (tratamientos) tuvo que ser formada con bloques de músculos cardiacos. Esto es, el tejido para un músculo cardiaco fue considerado como un bloque y se hicieron comparaciones de los tres tratamientos dentro de un músculo determinado. La ingesta de calcio para los tres niveles de digitalina, A, B y C, se compararon con base en los músculos del corazón de cuatro perros. Los resultados se muestran en la siguiente tabla.

Perros			
1	2	3	4
A 1342	C 1698	B 1296	A 1150
B 1608	B 1387	A 1029	C 1579
C 1881	A 1140	C 1549	B 1319

a Calcule las sumas de cuadrados para este experimento y construya una tabla ANOVA.

b ¿Cuántos grados de libertad están asociados con la SSE?

c ¿Los datos presentan suficiente evidencia para indicar una diferencia en la ingesta media de calcio para los tres niveles de digitalina?

d ¿Los datos indican una diferencia en la ingesta media de calcio para los músculos cardiacos de los cuatro perros?

e Indique la desviación estándar de la diferencia entre las ingestiones medias de calcio para dos niveles de digitalina.

f Encuentre un intervalo de confianza de 95% para la diferencia en las respuestas medias entre los tratamientos A y B.

13.84 Consulte el Ejercicio 13.83. ¿Aproximadamente cuántas repeticiones se necesitan para cada nivel de digitalina (cuántos bloques) para que el error de estimar la diferencia en la respuesta media para un par de niveles de digitalina sea menor que 20, con probabilidad .95? Suponga que se harían observaciones adicionales dentro de un diseño de bloques aleatorizado.

13.85 Se realizó un diseño completamente aleatorizado para comparar los efectos de cinco estímulos en los tiempo de reacción. Se emplearon 27 personas en el experimento, que se realizó usando un diseño completamente aleatorizado. Sin considerar los resultados del ANOVA, se desea comparar los estímulos A y D. Los tiempos de reacción (en segundos) son como se muestra en la siguiente tabla.

	Estímulo				
	A	B	C	D	E
	.8	.7	1.2	1.0	.6
	.6	.8	1.0	.9	.4
	.6	.5	.9	.9	.4
	.5	.5	1.2	1.1	.7
		.6	1.3	.7	.3
		.9	.8		
		.7			
Total	2.5	4.7	6.4	4.6	2.4
Media	.625	.671	1.067	.920	.480

a Realice un ANOVA y pruebe para una diferencia en los tiempos medios de reacción debida a los cinco estímulos. Indique límites para el valor p.

b Compare los estímulos A y D para ver si hay una diferencia en los tiempos medios de reacción. ¿Qué se puede decir acerca del nivel de significancia alcanzado?

13.86 Debido a que esperaríamos que el tiempo medio de reacción varíe de una persona a otra, el experimento del Ejercicio 13.85 pudo haberse realizado en forma más eficaz con el uso de un diseño de bloques aleatorizado con personas como bloques. En consecuencia, se emplearon cuatro personas en un nuevo experimento y cada persona fue sometida a cada uno de los cinco estímulos en orden aleatorio. Los tiempos de reacción (en segundos) son los mostrados en la siguiente tabla. Realice un ANOVA y pruebe las diferencias en los tiempos medios de reacción para los cuatro estímulos.

	Estímulo				
Sujeto	A	B	C	D	E
1	.7	.8	1.0	1.0	.5
2	.6	.6	1.1	1.0	.6
3	.9	1.0	1.2	1.1	.6
4	.6	.8	.9	1.0	.4

13.87 Consulte el Ejercicio 13.46. Construya intervalos de confianza para comparar cada una de las variedades de césped con el supremo Marvelgreen, en forma tal que el coeficiente simultáneo de confianza sea al menos .95. Interprete los resultados.

13.88 Demuestre que

$$\text{SS total} = \text{SST} + \text{SSB} + \text{SSE}$$

para un diseño de bloques aleatorizado, donde

$$\text{SSE} = \sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{ij} - \overline{Y}_{\bullet j} - \overline{Y}_{i\bullet} + \overline{Y})^2.$$

***13.89** Considere el siguiente modelo para las respuestas medidas en un diseño de bloques aleatorizado que contiene b bloques y k tratamientos:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

donde Y_{ij} = respuesta al tratamiento i en el bloque j,
μ = media general,
τ_i = efecto no aleatorio del tratamiento i, donde $\sum_{i=1}^{k}\tau_i = 0$,
β_j = efecto aleatorio del bloque j, donde las β_j son variables aleatorias independientes, distribuidas normalmente, con $E(\beta_j) = 0$ y $V(\beta_j) = \sigma_\beta^2$, para $j = 1, 2, \ldots, b$.
ε_{ij} = términos aleatorios de error donde las ε_{ij} son variables aleatorias independientes, distribuidas normalmente, con $E(\varepsilon_{ij}) = 0$ y $V(\varepsilon_{ij}) = \sigma_\varepsilon^2$, para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, b$.

Además, suponga que las β_j y las ε_{ij} también son independientes. Este modelo difiere del presentado en la Sección 13.8 en que se supone que los efectos de bloque son *variables aleatorias* en lugar de constantes fijas pero desconocidas.

a Si el modelo que acabamos de describir es apropiado, demuestre que las observaciones tomadas de bloques diferentes son independientes entre sí. Esto es, demuestre que Y_{ij} y $Y_{ij'}$ son independientes si $j \neq j'$, como son Y_{ij} y $Y_{i'j'}$ si $i \neq i'$ y $j \neq j'$.

b De acuerdo con el modelo que acabamos de describir, deduzca la covarianza de dos observaciones del mismo bloque. Esto es, encuentre $\text{Cov}(Y_{ij}, Y_{i'j})$ si $i \neq i'$.

c Dos variables aleatorias que tienen una distribución normal conjunta son independientes si y sólo si la covarianza de ellas es 0. Use el resultado del inciso b para determinar las condiciones en las que dos observaciones del mismo bloque son independientes entre sí.

***13.90** Consulte el modelo para el diseño de bloques aleatorizado con efecto de bloque al azar dado en el Ejercicio 13.89.

a Indique el valor y la varianza esperados de Y_{ij}.

b Con $\overline{Y}_{i\bullet}$ denote el promedio de todas las respuestas al tratamiento i. Use el modelo para el diseño de bloques aleatorizado para deducir $E(\overline{Y}_{i\bullet})$ y $V(\overline{Y}_{i\bullet})$. ¿$\overline{Y}_{i\bullet}$ es un estimador insesgado para la respuesta media al tratamiento i? ¿Por qué sí o por qué no? Observe que $V(\overline{Y}_{i\bullet})$ depende de b y de σ_β^2 y σ_ε^2.

c Considere $\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ para $i \neq i'$. Demuestre que $E(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = \tau_i - \tau_{i'}$. Este resultado implica que $\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ es un estimador insesgado de la diferencia en los efectos de los tratamientos i e i'.

d Deduzca $V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet})$. Observe que $V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet})$ depende sólo de b y σ_ε^2.

***13.91** Consulte el modelo para el diseño de bloques aleatorizado con efecto de bloque al azar dado en el Ejercicio 13.89 y denote con $\overline{Y}_{\bullet j}$ el promedio de todas las respuestas del bloque j. Deduzca

a $E(\overline{Y}_{\bullet j})$ y $V(\overline{Y}_{\bullet j})$.

b $E(\text{MST})$.

c $E(\text{MSB})$.

d $E(\text{MSE})$.

***13.92** Consulte el modelo para el diseño de bloques aleatorizado con efecto de bloque al azar dado en el Ejercicio 13.89 y los resultados obtenidos en el Ejercicio 13.91(c) y (d). Dé un estimador insesgado para

a σ_ε^2.

b σ_β^2.

***13.93** Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución normal con media μ y varianza σ^2. La independencia de $\sum_{i=1}^n (Y_i - \overline{Y})^2$ y $\overline{Y}$ se puede demostrar como sigue. Defina una matriz $\mathbf{A}$ de $n \times n$ con

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{-2}{\sqrt{2\cdot 3}} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & & \cdots & & \frac{1}{\sqrt{(n-1)n}} & \frac{-(n-1)}{\sqrt{(n-1)n}} \end{bmatrix}$$

y observe que $\mathbf{A}'\mathbf{A} = \mathbf{I}$, la matriz identidad. Entonces

$$\sum_{i=1}^n Y_i^2 = \mathbf{Y}'\mathbf{Y} = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y},$$

donde $\mathbf{Y}$ es el vector de valores Y_i.

a Demuestre que

$$\mathbf{AY} = \begin{bmatrix} \overline{Y}\sqrt{n} \\ U_1 \\ U_2 \\ \vdots \\ U_{n-1} \end{bmatrix},$$

donde $U_1, U_2, \ldots, U_{n-1}$ son funciones lineales de $Y_1, Y_2, \ldots, Y_n$. Entonces,

$$\sum_{i=1}^n Y_i^2 = n\overline{Y}^2 + \sum_{i=1}^{n-1} U_i^2.$$

b Demuestre que las funciones lineales $\overline{Y}\sqrt{n}, U_1, U_2, \ldots, U_{n-1}$ son ortogonales por pares y por tanto independientes de acuerdo con la suposición de normalidad. (Véase el Ejercicio 5.130.)

c Demuestre que

$$\sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^{n-1} U_i^2$$

y concluya que esta cantidad es independiente de $\overline{Y}$.

d Usando los resultados del inciso c, demuestre que

$$\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$$

tiene una distribución χ^2 con $(n - 1)$ grados de libertad.

***13.94** Considere un diseño de un factor con k tratamientos. Suponga que Y_{ij} es la j-ésima respuesta para el tratamiento (población) i y que Y_{ij} está distribuida normalmente con media μ_i y varianza σ^2, para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n_i$.

a Use el Ejercicio 13.93 para justificar que $\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_k$ son independientes de SSE.

b Demuestre que MST/MSE tiene una distribución F con $v_1 = k - 1$ y $v_2 = n_1 + n_2 + \cdots + n_k - k$ grados de libertad dada $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$. (Se puede suponer, para mayor sencillez, que $n_1 = n_2 = \cdots = n_k$.)

CAPÍTULO 14

Análisis de datos categóricos

14.1 Descripción del experimento

Numerosos experimentos resultan en mediciones que son *cualitativas* o *categóricas* más que *cuantitativas,* como muchas de las mediciones estudiadas en capítulos previos. En estos ejemplos una cualidad o característica es identificada por cada una de las unidades experimentales. Los datos asociados con esas mediciones se pueden resumir al dar la *cuenta* del número de mediciones que caen en cada una de las distintas categorías asociadas con la variable. Por ejemplo,

- Los empleados pueden ser clasificados en uno de cinco grupos de ingreso.
- Los ratones podrían reaccionar en una de tres formas cuando se someten a un estímulo.
- Los vehículos a motor podrían caer en uno de cuatro tipos de vehículos.
- Las pinturas podrían ser clasificadas en una de k categorías de acuerdo con su estilo y periodo.
- La calidad de incisiones quirúrgicas podrían ser más identificadas en forma significativa como excelente, muy buena, buena, regular o mala.
- Los artículos manufacturados son aceptables, de segunda o rechazados.

Con un grado razonable de aproximación, todos los ejemplos precedentes muestran las siguientes características que definen un *experimento multinomial* (vea la Sección 5.9):

1. El experimento consiste en n ensayos idénticos.
2. El resultado de cada ensayo cae en exactamente una de k categorías o celdas distintas.
3. La probabilidad de que el resultado de un solo ensayo caiga en una celda particular, la celda i, es p_i, donde $i = 1, 2, \ldots, k$ y continúa igual de un ensayo a otro. Observe que

$$p_1 + p_2 + p_3 + \cdots + p_k = 1.$$

4. Los ensayos son independientes.
5. Estamos interesados en $n_1, n_2, n_3, \ldots, n_k$, donde n_i para $i = 1, 2, \ldots, k$ es igual al número de ensayos para los cuales el resultado cae en la celda i. Observe que $n_1 + n_2 + n_3 + \cdots + n_k = n$.

Este experimento es análogo a lanzar n pelotas en k cajas, donde cada pelota debe caer exactamente en una de las cajas. La probabilidad de que una pelota caiga en una caja varía de una caja a otra pero sigue siendo la misma para cada caja en tiros repetidos. Por último, las pelotas son lanzadas en forma tal que los ensayos son independientes. Al término del experimento, observamos n_1 pelotas en la primera caja, n_2 en la segunda,… y n_k en la k-ésima. El número total de pelotas es $n = n_1 + n_2 + n_3 + \cdots + n_k$.

Observe la similitud entre los experimentos binomial y multinomial y, en particular, que el experimento binomial representa el caso especial para el experimento multinomial cuando $k = 2$. Las probabilidades de dos celdas, p y $q = 1 - p$, del experimento binomial son sustituidas por las probabilidades por celda k, $p_1, p_2, \ldots, p_k$, del experimento multinomial. El objetivo de este capítulo es hacer inferencias acerca de las probabilidades por celda $p_1, p_2, \ldots, p_k$. Las inferencias se expresarán en términos de pruebas estadísticas de hipótesis que se refieran a valores numéricos específicos de las probabilidades por celda o a su relación mutua.

Debido a que el cálculo de probabilidades multinomiales es un tanto engorroso, sería difícil calcular los niveles exactos de significancia (probabilidades de cometer errores tipo I) para las hipótesis que se refieren a los valores de $p_1, p_2, \ldots, p_k$. Por fortuna, hemos sido aliviados de este trabajo por el estadístico inglés Karl Pearson, quien propuso un estadístico muy útil para probar hipótesis con respecto a $p_1, p_2, \ldots, p_k$ y proporcionó la distribución muestral aproximada de este estadístico. En la siguiente sección haremos un resumen de la construcción del estadístico de Pearson.

14.2 Prueba ji cuadrada

Suponga que $n = 100$ pelotas fueron lanzadas a las celdas (cajas) y que sabíamos que p_1 era igual a 0.1. ¿Cuántas pelotas se esperaría que cayeran en la celda 1? Volviendo a la Sección 5.9, recuerde que n_1 tiene una distribución binomial (marginal) con parámetros n y p_1, y que

$$E(n_1) = np_1 = (100)(.1) = 10.$$

Del mismo modo, cada una de las n_i tiene distribuciones binomiales con parámetros n y p_i y el número esperado que caiga en la celda i es

$$E(n_i) = np_i, \qquad i = 1, 2, \ldots, k.$$

Ahora suponga que proponemos valores para $p_1, p_2, \dots, p_k$ y calculamos el valor esperado para cada celda. Ciertamente, si nuestra hipótesis es verdadera, los conteos n_i por celda no deben desviarse demasiado de sus valores esperados np_i para $i = 1, 2, \dots, k$. Por tanto, parecería intuitivamente razonable usar un estadístico de prueba que comprenda las k desviaciones,

$$n_i - E(n_i) = n_i - np_i, \qquad \text{para } i = 1, 2, \dots, k.$$

En 1900, Karl Pearson propuso el siguiente estadístico de prueba, que es una función de los cuadrados de las desviaciones de las cantidades observadas respecto de sus valores esperados, ponderados por los recíprocos de sus valores esperados:

$$X^2 = \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \sum_{i=1}^{k} \frac{[n_i - np_i]^2}{np_i}.$$

Aunque la prueba matemática está fuera del propósito de este texto, se puede demostrar que cuando n es grande, X^2 tiene una distribución de probabilidad Ji cuadrada (χ^2) aproximada. Podemos fácilmente demostrar este resultado para el caso $k = 2$ de la siguiente manera. Si $k = 2$, entonces $n_2 = n - n_1$ y $p_1 + p_2 = 1$. Entonces,

$$\begin{aligned}
X^2 &= \sum_{i=1}^{2} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} \\
&= \frac{(n_1 - np_1)^2}{np_1} + \frac{[(n - n_1) - n(1 - p_1)]^2}{n(1 - p_1)} \\
&= \frac{(n_1 - np_1)^2}{np_1} + \frac{(-n_1 + np_1)^2}{n(1 - p_1)} \\
&= (n_1 - np_1)^2 \left(\frac{1}{np_1} + \frac{1}{n(1 - p_1)} \right) = \frac{(n_1 - np_1)^2}{np_1(1 - p_1)}.
\end{aligned}$$

Hemos visto (Sección 7.5) que para n grande

$$\frac{n_1 - np_1}{\sqrt{np_1(1 - p_1)}}$$

tiene aproximadamente una distribución normal estándar. Como el cuadrado de una variable aleatoria normal estándar tiene una distribución χ^2 (vea Ejemplo 6.11), para $k = 2$ y n grande, X^2 tiene una distribución χ^2 aproximada con 1 grado de libertad (gl).

La experiencia ha demostrado que las cantidades n_i en la celda no deben ser tan pequeñas si la distribución χ^2 debe proporcionar una aproximación adecuada para la distribución X^2. Como regla práctica, requeriremos que todas las cantidades esperadas por celda sean al menos cinco, aunque Cochran (1952) ha observado que este valor puede ser de sólo uno para algunas situaciones.

Usted recordará el uso de la distribución de probabilidad χ^2 para probar una hipótesis respecto a una varianza poblacional σ^2 en la Sección 10.9. En particular, hemos visto que la forma de la distribución χ^2 y los cuantiles y áreas de cola asociados difieren en forma considerable dependiendo del número de grados de libertad (vea Tabla 6, Apéndice 3). Por tanto, si deseamos usar X^2 como estadístico de prueba, debemos conocer el número de grados de libertad asociado con la distribución χ^2 de aproximación y si hay que usar una prueba de una cola o de dos colas para localizar la región de rechazo para la prueba. Este último problema puede resolverse directamente. Como grandes diferencias entre las cantidades observadas y esperadas por celda contradicen la hipótesis nula, rechazaremos la hipótesis nula cuando X^2 sea grande y emplearemos una prueba estadística de cola superior.

La determinación del número apropiado de grados de libertad a emplearse para la prueba puede ser un poco complicada y, por tanto, será especificada para las aplicaciones físicas descritas en las secciones siguientes. Además, expresaremos el principio involucrado (que es fundamental para la prueba matemática de la aproximación) para que usted entienda por qué el número de grados de libertad cambia con varias aplicaciones. Este principio indica que *el número apropiado de grados de libertad será igual al número de celdas, k, menos* 1 *grado de libertad para cada restricción lineal independiente colocada en las probabilidades por celda.* Por ejemplo, una restricción lineal está *siempre* presente porque la suma de las probabilidades por celda debe ser igual a 1; esto es,

$$p_1 + p_2 + p_3 + \cdots + p_k = 1.$$

Otras restricciones se introducirán para algunas aplicaciones por la necesidad de estimar parámetros desconocidos requeridos en los cálculos de las frecuencias esperadas por celda, o por el método empleado para recolectar la muestra. Cuando deban estimarse parámetros desconocidos para calcular X^2, debe emplearse un estimador de máxima probabilidad (MLE). Los grados de libertad para la distribución χ^2 de aproximación se reduce en 1 para cada parámetro estimado. Estos casos aparecerán cuando consideremos varios ejemplos prácticos.

14.3 Prueba de una hipótesis con respecto a probabilidades especificadas por celda: una prueba de la bondad de ajuste

La hipótesis más sencilla respecto a probabilidades por celda es aquella que especifica valores numéricos para cada una. En este caso, estamos probando $H_0: p_1 = p_{1,0}, p_2 = p_{2,0}, \ldots, p_k = p_{k,0}$, donde $p_{i,0}$ denota un valor especificado para p_i. La hipótesis alternativa expresa, en general, que al menos una de las igualdades no se cumple. Como la única restricción en las probabilidades por celda es que $\sum_{i=1}^{k} p_i = 1$, el estadístico de prueba X^2 tiene aproximadamente una distribución χ^2 con $k - 1$ grados de libertad.

EJEMPLO **14.1** Un grupo de ratas, una por una, bajan por una rampa que conduce a tres puertas. Deseamos probar la hipótesis de que las ratas no tienen preferencia respecto a la elección de una puerta. Entonces, la hipótesis nula apropiada es

$$H_0: p_1 = p_2 = p_3 = \frac{1}{3},$$

donde p_i es la probabilidad de que una rata escogerá la puerta i, para $i = 1, 2$ o 3.

Suponga que las ratas bajaron por la rampa $n = 90$ veces y que las tres frecuencias por celda observadas fueron $n_1 = 23$, $n_2 = 36$ y $n_3 = 31$. La frecuencia esperada por celda es igual para cada celda: $E(n_i) = np_i = (90)(1/3) = 30$. Las frecuencias por celda observada y espe-

Tabla 14.1 Cantidades por celda observadas y esperadas

	Puerta		
Valor	1	2	3
Frecuencia por celda observada	$n_1 = 23$	$n_2 = 36$	$n_3 = 31$
Frecuencia por celda esperada	(30)	(30)	(30)

rada se presentan en la Tabla 14.1. Observe la discrepancia entre las frecuencias observadas y esperadas por celda. ¿Los datos presentan suficiente evidencia para justificar el rechazo de la hipótesis de no preferencia?

Solución El estadístico de prueba χ^2 para nuestro ejemplo tendrá $(k - 1) = 2$ grados de libertad porque la única restricción en las probabilidades por celda es que

$$p_1 + p_2 + p_3 = 1.$$

Por tanto, si escogemos $\alpha = .05$, rechazaríamos la hipótesis nula cuando $X^2 > 5.991$ (vea la Tabla 6, Apéndice 3).

Sustituyendo en la fórmula para X^2 obtenemos

$$\begin{aligned} X^2 &= \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \\ &= \frac{(23-30)^2}{30} + \frac{(36-30)^2}{30} + \frac{(31-30)^2}{30} = 2.87. \end{aligned}$$

Como X^2 es menor que el valor crítico tabulado de χ^2, la hipótesis nula no se rechaza y concluimos que los datos no presentan suficiente evidencia para indicar que las ratas tienen preferencia por cualquiera de las puertas. En este caso, el valor p está dado por valor $p = P(\chi^2 > 2.87)$, donde χ^2 posee una distribución χ^2 con $k - 1 = 2$ grados de libertad. Usando la Tabla 6, Apéndice 3, se deduce que el valor $p > 0.10$. La aplicación *Chi-Square Probability and Quantiles* da un valor $p = P(\chi^2 > 2.87) = .23812$. ■

El estadístico χ^2 también se puede usar para probar si los datos muestrales indican que un modelo específico para una distribución poblacional no se ajusta a los datos. Un ejemplo de esa prueba, llamada *prueba de bondad de ajuste*, se ilustra en el siguiente ejemplo.

EJEMPLO **14.2** El número de accidentes Y por semana en un crucero se verificó por $n = 50$ semanas, con los resultados que se muestran en la Tabla 14.2. Pruebe la hipótesis de que la variable aleatoria Y tiene una distribución de Poisson, suponiendo que las observaciones son independientes. Use $\alpha = .05$.

Solución La hipótesis nula H_0 expresa que Y tiene la distribución de Poisson dada por

$$p(y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \qquad y = 0, 1, 2, \ldots.$$

Tabla 14.2 Datos para el Ejemplo 14.2

y	Frecuencia
0	32
1	12
2	6
3 o más	0

Como λ es desconocida, debemos hallar su estimador de máxima probabilidad (MLE). En el Ejercicio 9.80 establecimos que el MLE de λ es $\hat{\lambda} = \overline{Y}$. Para los datos dados, $\hat{\lambda}$ tiene el valor $\overline{y} = 24/50 = .48$.

Tenemos, para los datos dados, tres celdas con cinco o más observaciones: las celdas definidas por $Y = 0$, $Y = 1$ y $Y \geq 2$. Dada H_0, las probabilidades para estas celdas son

$$p_1 = P(Y = 0) = e^{-\lambda}, \qquad p_2 = P(Y = 1) = \lambda e^{-\lambda},$$

$$p_3 = P(Y \geq 2) = 1 - e^{-\lambda} - \lambda e^{-\lambda}.$$

Estas probabilidades se calculan al sustituir λ con $\hat{\lambda}$, que da como resultado

$$\hat{p}_1 = e^{-.48} = .619, \qquad \hat{p}_2 = .48e^{-.48} = .297, \qquad \hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2 = .084.$$

Si las observaciones son independientes, las frecuencias por celda n_1, n_2 y n_3 tienen una distribución multinomial con parámetros p_1, p_2 y p_3. Entonces, $E(n_i) = np_i$ y las frecuencias por celda esperadas y estimadas están dadas por

$$\widehat{E(n_1)} = n\hat{p}_1 = 30.95, \qquad \widehat{E(n_2)} = n\hat{p}_2 = 14.85, \qquad \widehat{E(n_3)} = n\hat{p}_3 = 4.20.$$

Entonces, el estadístico de prueba está dado por

$$X^2 = \sum_{i=1}^{3} \frac{[n_i - \widehat{E(n_i)}]^2}{\widehat{E(n_i)}},$$

que tiene aproximadamente una distribución χ^2 con $(k - 2) = 1$ grado de libertad. (Un grado de libertad se pierde porque λ tenía que calcularse, el otro, porque $\sum_{i=1}^{3} p_i = 1$.)

Al calcular X^2 encontramos

$$X^2 = \frac{(32 - 30.95)^2}{30.95} + \frac{(12 - 14.85)^2}{14.85} + \frac{(6 - 4.20)^2}{4.20} = 1.354.$$

Como $\chi^2_{.05} = 3.841$, con 1 grado de libertad, no rechazamos H_0. Los datos no presentan suficiente evidencia para contradecir nuestra hipótesis de que Y posee una distribución de Poisson. El valor p está dado por $P(\chi^2 > 1.354)$. La Tabla 6, Apéndice 3, da el valor $p > .10$ mientras que la aplicación *Chi-Square Probability and Quantiles* establece que el valor $p = .24458$. A menos que se utilice un valor muy grande de ($\alpha \geq .24458$), no hay suficiente evidencia para rechazar la afirmación de que el número de accidentes por semana tiene una distribución de Poisson. ■

Ejercicios

14.1 Históricamente, las proporciones de todas las personas de origen caucásico en Estados Unidos con fenotipos sanguíneos A, B, AB y O son .41, .10, .04 y .45, respectivamente. Para determinar si las proporciones actuales de población todavía se comparan con estos valores históricos, se seleccionó una muestra aleatoria de 200 estadounidenses caucásicos y se registraron sus fenotipos sanguíneos. Los números observados con cada fenotipo se dan en la siguiente tabla.

A	B	AB	O
89	18	12	81

a ¿Hay suficiente evidencia, en el nivel de significancia .05 para afirmar que las proporciones actuales difieren de los valores históricos?

b **Ejercicio Applet** Use la aplicación *Chi-Square Probability and Quantiles* para hallar el valor p asociado con la prueba del inciso a.

14.2 Registros previos de inscripciones en una gran universidad indican que del número total de personas que solicitan admisión, 60% son admitidos incondicionalmente, 5% son admitidos condicionalmente y el resto son rechazados. De 500 solicitantes para el año siguiente, 329 fueron admitidos incondicionalmente, 43 fueron admitidos condicionalmente y el resto no fueron admitidos. ¿Estos datos indican una desviación de los porcentajes previos de admisión?

a Pruebe usando $\alpha = .05$.

b **Ejercicio Applet** Use la aplicación *Chi-Square Probability and Quantiles* para hallar el valor p asociado con la prueba del inciso a.

14.3 Una vía rápida en una ciudad, con cuatro carriles en cada dirección, fue estudiada para ver si los conductores prefieren viajar en los carriles interiores. Se observaron un total de 1000 automóviles durante el intenso tránsito de la hora pico por la mañana y se registraron sus carriles respectivos. Los resultados se muestran en la siguiente tabla. ¿Los datos presentan suficiente evidencia para indicar que algunos carriles se prefieren a otros? (Pruebe la hipótesis de que $p_1 = p_2 = p_3 = p_4 = 1/4$, usando $\alpha = .05$.) Proporcione límites para el valor p asociado.

Carril	1	2	3	4
Cantidad	294	276	238	192

14.4 ¿Odia usted los lunes? Investigadores en Alemania han dado otra razón para hacerlo: concluyeron que el riesgo de ataque al corazón en un lunes, para una persona que trabaja, puede ser hasta 50% mayor que en cualquier otro día.[1] Los investigadores registraron ataques al corazón y paros cardiacos en un periodo de 5 años entre 330 000 personas que vivían cerca de Augsberg, Alemania. En un intento por verificar lo dicho por el investigador, se encuestaron 200 trabajadores que habían tenido ataques al corazón recientemente. El día en el que ocurrieron sus ataques al corazón aparecen en la tabla siguiente.

Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
24	36	27	26	32	26	29

¿Estos datos presentan suficiente evidencia para indicar que hay una diferencia en los porcentajes de ataques al corazón que ocurren en diferentes días de la semana? Pruebe usando $\alpha = .05$.

14.5 Después de inspeccionar los datos del Ejercicio 14.4, quizá le gustaría probar la hipótesis de que la probabilidad de que una persona sufra un ataque al corazón en lunes es 1/7 contra la alternativa de que esta probabilidad sea mayor que 1/7.

1. *Fuente:* Daniel Q. Haney, "Mondays May Be Hazardous," *Press-Enterprise* (Riverside, Calif.), 17 de noviembre de 1992, p. A16.

a Realice la prueba anterior usando $\alpha = .05$.

b ¿Qué principio de una buena práctica estadística se viola en la prueba del inciso a?

c Antes de ver los datos actuales, ¿hay una razón que por la cual usted padría considerar legítimamente las hipótesis del inciso a?

14.6 Imagine que las suposiciones asociadas con un experimento multinomial están todas satisfechas. Entonces (vea la Sección 5.9) cada una de las n_i, $i = 1, 2, \ldots, k$, tienen una distribución binomial con parámetros n y p_i. Además, Cov $(n_i, n_j) = -np_i p_j$ si $i \neq j$.

a ¿Cuál es $E(n_i - n_j)$?

b Consulte el inciso a. Proporcione un estimador insesgado para $p_i - p_j$.

c Demuestre que $V(n_i - n_j) = n[p_i(1 - p_i) + p_j(1 - p_j) + 2p_i p_j]$.

d Consulte el inciso c. ¿Cuál es la varianza del estimador insesgado que se dio en el inciso b?

e Proporcione un estimador consistente para $n^{-1}V(n_i - n_j)$.

f Si n es grande, el estimador que se dio en el inciso b está distribuido normalmente en forma aproximada con media $p_i - p_j$ y varianza $n^{-2}V(n_i - n_j)$. Si $\hat{p}_i = n_i/n$ y $\hat{p}_j = n_j/n$, demuestre que un intervalo de confianza $(1 - \alpha)100\%$ con muestra grande para $p_i - p_j$ está dado por

$$\hat{p}_i - \hat{p}_j \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_i(1 - \hat{p}_i) + \hat{p}_j(1 - \hat{p}_j) + 2\hat{p}_i\hat{p}_j}{n}}.$$

14.7 Consulte el Ejercicio 14.3. El carril 1 es para vehículos “lentos” y el carril 4 es para vehículos “rápidos”. Use la fórmula de intervalo de confianza dada en el Ejercicio 14.6(f) para dar un intervalo de confianza de 95% para $p_1 - p_4$. ¿Se concluiría que una mayor proporción viaja en el carril lento que en el carril rápido? ¿Por qué?

14.8 La teoría de Mendel dice que el número de un tipo de chícharos que cae en las clasificaciones redonda y amarilla, arrugada y amarilla, redonda y verde, y arrugada y verde debe estar en la proporción 9:3:3:1. Suponga que 100 de estos chícharos revelaron 56, 19, 17 y 8 en las respectivas categorías. ¿Estos datos son consistentes con el modelo? Use $\alpha = .05$. (La expresión 9:3:3:1 significa que $9/16$ de los chícharos deben ser redondos y amarillos, 3/16 deben ser arrugados y amarillos, etc.)

14.9 Consulte el Ejercicio 14.6(f) y los datos del Ejercicio 14.8.

a Proporcione un intervalo de confianza de 95% para la diferencia en las proporciones de chícharos redondos amarillos y redondos verdes.

b Construya, usando el método Bonferroni estudiado en la Sección 13.12, intervalos simultáneos de confianza para comparar la proporción de chícharos redondos amarillos con las proporciones de chícharos en cada una de las otras tres categorías. Los intervalos han de tener coeficiente simultáneo de confianza de al menos .95.

14.10 Dos tipos de defectos, A y B, se ven con frecuencia a la salida de un proceso de manufactura. Cada artículo puede ser clasificado en una de las cuatro clases: $A \cap B$, $A \cap \overline{B}$, $\overline{A} \cap B$ y $\overline{A} \cap \overline{B}$, donde $\overline{A}$ denota la ausencia del defecto tipo A. Para 100 artículos inspeccionados, se observaron las siguientes frecuencias:

$$A \cap B : 48, \qquad A \cap \overline{B} : 18, \qquad \overline{A} \cap B : 21, \qquad \overline{A} \cap \overline{B} : 13.$$

¿Hay suficiente evidencia para indicar que las cuatro categorías, en el orden citado, no se presentan en la proporción 5:2:2:1? (Use $\alpha = .05$.)

14.11 Los datos de la tabla siguiente son los conteos de frecuencia para 400 observaciones del número de colonias bacterianas dentro del campo de un microscopio, usando muestras de película de leche.[2] ¿Hay suficiente evidencia para decir que los datos no se ajustan a la distribución de Poisson? (Use $\alpha = .05$.)

2. *Fuente:* C. A. Bliss y R. A. Fisher, “Fitting the Negative Binomial Distribution to Biological Data,” *Biometrics* 9 (1953): 176-200. Biometrics Society. Todos los derechos reservados.

Número de colonias por campo	Frecuencia de observación
0	56
1	104
2	80
3	62
4	42
5	27
6	9
7	9
8	5
9	3
10	2
11	0
19	1
	400

14.12 Durante un periodo fijo se observó el número de accidentes sufridos por mecánicos, con los resultados que se ven en la siguiente tabla.[3] Pruebe, con un nivel de significancia de 5%, la hipótesis de que los datos provienen de una distribución de Poisson.

Accidentes por mecánico	Frecuencia de observación (número de mecánicos)
0	296
1	74
2	26
3	8
4	4
5	4
6	1
7	0
8	1

14.4 Tablas de contingencia

Un problema frecuente en el análisis de conteo de datos se refiere a la evaluación de la independencia de dos métodos para la clasificación de sujetos (personas). Por ejemplo, podríamos clasificar una muestra de personas por género y por opinión acerca de un problema político, para probar la hipótesis de que las opiniones sobre el problema son independientes del género. Del mismo modo, podríamos clasificar pacientes que sufren de una enfermedad de acuerdo con el tipo de medicamento y su rapidez de recuperación, para ver si ésta depende del tipo de medicamento. En cada uno de estos ejemplos deseamos investigar la *dependencia* (o *contingencia*) entre dos criterios de clasificación.

Suponga que deseamos clasificar defectos hallados en muebles producidos en una planta manufacturera de acuerdo con (1) el tipo del defecto y (2) el turno de producción. Se registró

3. *Fuente:* C. A. Bliss y R. A. Fisher, "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics* 9 (1953): 176-200. Biometrics Society. Todos los derechos reservados.

Tabla 14.3 Tabla de contingencia

	Tipo de defecto				
Turno	A	B	C	D	Total
1	15 (22.51)	21 (20.99)	45 (38.94)	13 (11.56)	94
2	26 (22.99)	31 (21.44)	34 (39.77)	5 (11.81)	96
3	33 (28.50)	17 (26.57)	49 (49.29)	20 (14.63)	119
Total	74	69	128	38	309

un total de $n = 309$ defectos en muebles y los defectos se clasificaron como uno de cuatro tipos, A, B, C o D. Al mismo tiempo, cada mueble fue identificado de acuerdo con el turno de producción durante el cual fue manufacturado. Estas cantidades se presentan en la Tabla 14.3, ejemplo de *tabla de contingencia.* (Como veremos más adelante, los números dentro de paréntesis son las frecuencias esperadas y estimadas por celda.) Nuestro objetivo es probar la hipótesis nula de que el tipo de defecto es independiente del turno contra la alternativa de que los dos esquemas de clasificación son dependientes. Esto es, deseamos probar H_0 : clasificación de columna es independiente de clasificación de renglón.

Sea p_{A} igual a la probabilidad incondicional de que un defecto sea del tipo A. De manera similar, defina p_{B}, p_{C} y p_{D} como las probabilidades de observar los otros tres tipos de defectos. Entonces estas probabilidades, a las que llamaremos *probabilidades de columna* de la Tabla 14.3, satisfacen el requisito de que

$$p_{\text{A}} + p_{\text{B}} + p_C + p_{\text{D}} = 1.$$

Del mismo modo, sea p_i para $i = 1$, 2 o 3 igual a las *probabilidades de renglón* de que una pieza defectuosa se produzca en el turno i, donde

$$p_1 + p_2 + p_3 = 1.$$

Si las dos clasificaciones son independientes entre sí, cada probabilidad por celda es igual al producto de sus respectivas probabilidades de renglón y columna. Por ejemplo, la probabilidad de que un defecto se presente en el turno 1 y sea del tipo A es $p_1 \times p_{\text{A}}$. Observamos que los valores numéricos de las probabilidades por celda no están especificados en el problema en consideración. La hipótesis nula sólo especifica que cada probabilidad por celda es igual al producto de sus respectivas probabilidades de renglón y columna y, por tanto, implica independencia de las dos clasificaciones.

El análisis de los datos obtenidos de una tabla de contingencia difiere del análisis del Ejemplo 14.1, porque debemos *estimar* las probabilidades de renglón y columna para estimar las frecuencias por celda esperadas. Las frecuencias por celda esperadas y estimadas pueden ser sustituidas por la $E(n_i)$ en X^2, y X^2 continuará teniendo una distribución que está bien calculada por una distribución de probabilidad χ^2.

El estimador de máxima probabilidad (MLE) para cualquier probabilidad de renglón o columna se encuentra como sigue. Con n_{ij} denotemos la frecuencia observada en el renglón i y la columna j de la tabla de contingencia y con p_{ij} denotemos la probabilidad de que una observación caiga en esta celda. Si las observaciones se seleccionan de manera independiente, entonces las frecuencias por celda tienen una distribución multinomial y el MLE de p_{ij} es simplemente la frecuencia relativa observada para esa celda. Esto es,

$$\hat{p}_{ij} = \frac{n_{ij}}{n}, \qquad i = 1, 2, \ldots, r,\ j = 1, 2, \ldots, c$$

(véase el Ejercicio 9.87).

Del mismo modo, viendo el renglón i como una sola celda, la probabilidad para el renglón i está dada por p_i, y si r_i denota el número de observaciones en el renglón i,

$$\hat{p}_i = \frac{r_i}{n}$$

es el MLE de p_i.

Por argumentos semejantes, el MLE de la probabilidad de la j-ésima columna es c_j/n, donde c_j denota el número de observaciones en la columna j.

Dada la hipótesis nula, el MLE del valor esperado de n_{11} es

$$\widehat{E(n_{11})} = n(\hat{p}_1 \times \hat{p}_A) = n\left(\frac{r_1}{n}\right)\left(\frac{c_1}{n}\right) = \frac{r_1 \cdot c_1}{n}.$$

De manera análoga, si la hipótesis nula es verdadera, el valor esperado y estimado de la frecuencia por celda, n_{ij} para una tabla de contingencia es igual al producto de sus respectivos totales de renglón y columna divididos entre el tamaño muestral total. Esto es,

$$\widehat{E(n_{ij})} = \frac{r_i c_j}{n}.$$

Las frecuencias esperadas y estimadas por celda para nuestro ejemplo se muestran entre paréntesis en la Tabla 14.3. Por ejemplo,

$$\widehat{E(n_{11})} = \frac{r_1 c_1}{n} = \frac{94(74)}{309} = 22.51.$$

Ahora podemos usar las frecuencias esperadas y observadas por celda que se muestran en la Tabla 14.3 para calcular el valor del estadístico de prueba

$$\begin{aligned} X^2 &= \sum_{j=1}^{4}\sum_{i=1}^{3} \frac{[n_{ij} - \widehat{E(n_{ij})}]^2}{\widehat{E(n_{ij})}} \\ &= \frac{(15 - 22.51)^2}{22.51} + \frac{(26 - 22.99)^2}{22.99} + \cdots + \frac{(20 - 14.63)^2}{14.63} \\ &= 19.17. \end{aligned}$$

El único obstáculo restante comprende la determinación del número apropiado de grados de libertad asociado con el estadístico de prueba. Daremos éste como regla, que más adelante justificamos. *Los grados de libertad asociados con una tabla de contingencia que tenga r renglones y c columnas siempre será igual a* $(r - 1)(c - 1)$. Para nuestro ejemplo, comparamos X^2 con el valor crítico de χ^2 con $(r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$ grados de libertad.

Usted recordará que el número de grados de libertad asociado con el estadístico χ^2 será igual al número de celdas (en este caso, $k = r \times c$) menos 1 grado de libertad por cada restricción lineal independiente colocada en las probabilidades por celda. El número total de celdas para los datos de la Tabla 14.3 es $k = 12$. De esta cantidad restamos 1 grado de libertad porque la suma de las probabilidades por celda debe ser igual a 1; esto es,

$$p_{11} + p_{12} + \cdots + p_{34} = 1.$$

Además, usamos las frecuencias por celda para estimar dos de las tres probabilidades de renglón. Observe que la estimación de la probabilidad del tercer renglón se determina una vez que hayamos estimado p_1 y p_2, porque

$$p_1 + p_2 + p_3 = 1.$$

Entonces, perdemos $3 - 1 = 2$ grados de libertad para estimar las probabilidades de renglón.

Por último, usamos las frecuencias de las celdas para estimar $(c - 1) = 3$ probabilidades de columna y por tanto perdemos $(c - 1) = 3$ grados de libertad adicionales. El número total de grados de libertad restante es

$$\text{gl} = 12 - 1 - 2 - 3 = 6 = (3 - 1)(4 - 1).$$

En general vemos que el número total de grados de libertad asociado con una tabla de contingencia de $r \times c$ es

$$\text{gl} = rc - 1 - (r - 1) - (c - 1) = (r - 1)(c - 1).$$

Por tanto, en nuestro ejemplo que relaciona el turno con el tipo de defecto del mueble, si usamos $\alpha = .05$, rechazaremos la hipótesis nula de que las dos clasificaciones son independientes si $X^2 > 12.592$. Como el valor del estadístico de prueba, $X^2 = 19.17$, excede al valor crítico de χ^2, rechazamos la hipótesis nula en el nivel de significancia $\alpha = .05$. El valor p asociado está dado por valor $p = P(\chi^2 > 19.17)$. Los límites de esta probabilidad se pueden obtener usando la Tabla 6, Apéndice 3, de la cual se deduce que valor $p < .005$. La aplicación *Chi-Square Probability and Quantiles* da el valor p exacto $= .00389$. En consecuencia, para cualquier valor de α mayor o igual a .00389, los datos presentan suficiente evidencia para indicar dependencia entre el tipo de defecto y el turno de manufactura. Es probable que un estudio de las operaciones de producción para los tres turnos permita descubrir la causa.

EJEMPLO **14.3** Se realizó un estudio para evaluar la efectividad de una nueva vacuna para la gripe que tenía que ser administrada en una pequeña comunidad. La vacuna era gratuita para quienes la solicitaran, en una secuencia de dos inyecciones durante un periodo de 2 semanas; algunas personas recibieron esa secuencia, pero otras se presentaron sólo para la primera y los demás no recibieron ninguna.

Un estudio de 1000 habitantes de la localidad arrojó en la primavera siguiente la información que se indica en la Tabla 14.4. ¿Los datos presentan suficiente evidencia para indicar una dependencia entre las dos clasificaciones, es decir, categoría de vacuna y aparición o no aparición de la gripe?

Solución La pregunta investiga si los datos aportan suficiente evidencia para indicar dependencia entre categoría de vacuna y aparición o no aparición de la gripe. Por tanto, analizamos los datos como una tabla de contingencia.

Tabla 14.4 Tabulación de datos para el Ejemplo 14.3

Condición	Sin vacuna	Una vacuna	Dos vacunas	Total
Gripe	24 (14.4)	9 (5.0)	13 (26.6)	46
Sin gripe	289 (298.6)	100 (104.0)	565 (551.4)	954
Total	313	109	578	1000

Las frecuencias esperadas y estimadas por celda pueden calcularse con el uso de los totales apropiados de renglón y columna,

$$\widehat{E(n_{ij})} = \frac{r_i c_j}{n}.$$

Así, por ejemplo,

$$\widehat{E(n_{11})} = \frac{r_1 c_1}{n} = \frac{(46)(313)}{1000} = 14.4,$$
$$\widehat{E(n_{12})} = \frac{r_1 c_2}{n} = \frac{(46)(109)}{1000} = 5.0.$$

Éstas y las frecuencias esperadas y estimadas por celda restantes se muestran entre paréntesis en la Tabla 14.4.

El valor del estadístico de prueba X^2 se calcula y compara ahora con el valor crítico de χ^2 que posee $(r-1)(c-1) = (1)(2) = 2$ grados de libertad. Entonces, para $\alpha = .05$, rechazaremos la hipótesis nula cuando $X^2 > 5.991$. Sustituyendo en la fórmula para X^2, obtenemos

$$\begin{aligned} X^2 &= \frac{(24-14.4)^2}{14.4} + \frac{(289-298.6)^2}{298.6} + \cdots + \frac{(565-551.4)^2}{551.4} \\ &= 17.35. \end{aligned}$$

Observe que si X^2 cae en la región de rechazo, entonces rechazamos la hipótesis nula de independencia de las dos clasificaciones. Si elegimos usar el método de nivel de significancia alcanzado para hacer nuestra inferencia, el uso de la Tabla 6, Apéndice 3, establece que el valor $p < .005$. La aplicación χ^2 da un valor $p = .00017$. En este caso, como siempre, encontramos un acuerdo entre nuestro método de nivel α fijado para probar y la interpretación apropiada del valor p. ■

Como se estableció en la Sección 5.9, las n_{ij} están correlacionadas negativamente. Por ejemplo, $\text{Cov}(n_{ij}, n_{kl}) = -np_{ij}p_{kl}$ si $i \neq k$ o $j \neq l$. Una adaptación del resultado dado en el Ejercicio 14.7(f) se puede usar para asignar un intervalo de confianza con muestra grande para $p_{ij} - p_{kl}$ si dicho intervalo tiene un valor interpretativo práctico. Del mismo modo, las proporciones marginales se pueden comparar si se "colapsa" la tabla de contingencia a sólo las observaciones marginales de renglón o columna. El resultado del Ejercicio 14.6 (f) se aplica directamente a la tabla colapsada. No obstante, estas tablas marginales "colapsadas" sacrifican cualquier información acerca de la dependencia entre las variables de renglón y columna.

Hemos considerado sólo la hipótesis más sencilla conectada con una tabla de contingencia, la de independencia entre renglones y columnas. Muchas otras hipótesis son posibles y numerosas técnicas se han ideado para probarlas. Para más información sobre este tema, consulte la obra de Agresti (2002) y Fienberg (1980).

Ejercicios

14.13 En el aniversario 40 del asesinato del presidente John F. Kennedy, una encuesta del noticiero FOX mostró que casi todos los estadounidenses están en desacuerdo con las conclusiones del gobierno acerca de ese crimen. La Comisión Warren encontró que Lee Harvey Oswald actuó solo cuando le disparó a

Kennedy, pero muchos estadounidenses no están seguros de esta conclusión. ¿Piensa usted que sabemos todos los hechos relevantes asociados con el asesinato de Kennedy o piensa que se ha ocultado información? La siguiente tabla contiene los resultados de una encuesta nacional de 900 votantes registrados.[4]

	Conocemos todos los datos relevantes	Se ocultan algunos datos relevantes	No está seguro
Demócrata	42	309	31
Republicano	64	246	46
Otro	20	115	27

a ¿Los datos dan suficiente evidencia para indicar una dependencia entre afiliación de un partido y la opinión acerca de un posible encubrimiento? Pruebe usando $\alpha = .05$.

b Fije límites para el valor p asociado e interprete el resultado.

c **Ejercicio Applet** Use la aplicación χ^2 para obtener el valor p aproximado.

d ¿Por qué es "aproximado" el valor que obtuvo en el inciso c?

14.14 Joseph Jacobson y Diane Wille realizaron un estudio para determinar el efecto del cuidado temprano de niños con patrones de apego entre hijo y madre.[5] En el estudio, 93 infantes fueron clasificados como "seguro" o "ansioso" usando el paradigma Ainsworth de situación extraña. Además, los infantes fueron clasificados de acuerdo con el número promedio de horas por semana que recibían cuidado. Los datos aparecen en la siguiente tabla.

	Horas en cuidados de infantes		
Patrón de afecto	Bajo (0-3 horas)	Moderado (4-19 horas)	Alto (20-54 horas)
Seguro	24	35	5
Ansioso	11	10	8

a ¿Los datos indican dependencia entre patrones de apego y el número de horas de atención al niño? Pruebe usando $\alpha = .05$.

b Establezca límites para el nivel de significancia alcanzado.

14.15 Suponga que las entradas en una tabla de contingencia que aparecen en el renglón i y la columna j están denotadas por n_{ij}, para $i = 1, 2, \ldots, r$ y $j = 1, 2, \ldots, c$; que los totales de renglón y columna están denotados por r_i, para $i = 1, 2, \ldots, r$, y c_j, para $j = 1, 2, \ldots, c$; y que el tamaño muestral total es n.

a Demuestre que

$$X^2 = \sum_{j=1}^{c}\sum_{i=1}^{r} \frac{[n_{ij} - \widehat{E(n_{ij})}]^2}{\widehat{E(n_{ij})}} = n\left(\sum_{j=1}^{c}\sum_{i=1}^{r} \frac{n_{ij}^2}{r_i c_j} - 1\right).$$

Observe que esta fórmula proporciona una forma más eficiente desde el punto de vista computacional para calcular el valor de X^2.

b Usando la fórmula anterior, ¿qué le ocurre al valor de X^2 si toda entrada en la tabla de contingencia se multiplica por la misma constante entera $k > 0$?

14.16 Una encuesta para explorar la relación entre los patrones de asistencia a la iglesia de votantes y sus elecciones de candidato presidencial se publicó en *Riverside Press-Enterprise* antes de la elección de

4. *Fuente:* adaptado de Dana Blanton, "Poll: Most Believe 'Cover-up' of JFK Assassination Facts," http://www.foxnews.com/story/0,2933,102511,00.html, 10 de febrero de 2004.

5. *Fuente:* Linda Schmittroth (ed.) *Statistical Record of Women Worldwide* (Detroit and London: Gale Research, 1991), pp. 8, 9, 335.

2004. A los votantes se les preguntó con qué frecuencia asistían a la iglesia y por cuál de los dos candidatos presidenciales (George W. Bush o John Kerry) iban a votar en la elección. Los resultados de una encuesta similar están contenidos en la siguiente tabla.[6]

Asistencia a la iglesia	Bush	Kerry
Más de una vez por semana	89	53
Una vez por semana	87	68
Una o dos veces al mes	93	85
Una o dos veces al año	114	134
Rara vez/nunca	22	36

a ¿Hay evidencia suficiente para indicar dependencia entre la frecuencia de asistencia a la iglesia y la elección del candidato en la elección presidencial de 2004? Pruebe con un nivel de significancia de .05. Ponga límites al nivel de significancia alcanzado.

b Fije un intervalo de confianza de 95% para la proporción de personas que informan haber asistido a la iglesia al menos una vez por semana.

14.17 En el mundo académico es frecuente que estudiantes y asesores de profesores colaboren en publicaciones de investigación, produciendo obras en las que el crédito de la publicación puede tomar varias formas. Muchos piensan que la primera autoría de un ensayo técnico de un estudiante debe dársele a éste, a menos que la participación del asesor sea muy importante. En un intento por ver si de hecho éste es el caso, se estudió el crédito de autoría para diferentes grados de aportación del asesor y dos objetivos (disertaciones contra investigación sin título académico). La frecuencia de decisiones de asignación de autoría para disertaciones publicadas se da en las siguientes tablas, tal como fueron asignadas por 60 miembros del profesorado y 161 estudiantes:[7]

Profesores participantes

Asignación de autoría	Entrada elevada	Entrada media	Entrada baja
Asesor como primer autor, estudiante como segundo autor (obligatorio)	4	0	0
Estudiante como primer autor, asesor como segundo autor (obligatorio)	15	12	3
Estudiante como primer autor, asesor como segundo autor (cortesía)	2	7	7
Estudiante como único autor	2	3	5

Estudiantes participantes

Asignación de autoría	Entrada elevada	Entrada media	Entrada baja
Asesor como primer autor, estudiante como segundo autor (obligatorio)	19	6	2
Estudiante como primer autor, asesor como segundo autor (obligatorio)	19	41	27
Estudiante como primer autor, asesor como segundo autor (cortesía)	3	7	31
Estudiante como único autor	0	3	3

6. *Fuente*: Adaptado de Bettye Wells Miller, "Faith Shows Ballot Clout," *Press-Enterprise* (Riverside, Calif.), 1 de marzo de 2004, p. A7.

7. *Fuente*: M. Martin Costa y M. Gatz, "Determination of Authorship Credit in Publisher Dissertations," *Psychological Science* 3(6) (1992): 54.

a ¿Hay suficiente evidencia para indicar una dependencia entre la asignación de autoría y la entrada del asesor, según lo consideren los miembros del profesorado? Pruebe usando $\alpha = .01$.

b ¿Hay suficiente evidencia para indicar una dependencia entre la asignación de autoría y la entrada del asesor, según lo consideren los estudiantes? Pruebe usando $\alpha = .01$.

c ¿Han sido violadas algunas de las suposiciones necesarias para un análisis válido en los incisos a y b? ¿Qué efecto podría tener esto en la validez de sus conclusiones?

14.18 Un estudio de la cantidad de violencia en televisión en lo que respecta a la edad del televidente dio los resultados que se muestran en la siguiente tabla, para 81 personas. (Cada una de las personas del estudio fue clasificada, de acuerdo con los hábitos de ver TV de la persona, como que ve poca violencia o mucha violencia.) ¿Los datos indican que ver violencia no depende de la edad del televidente con un nivel de significancia de 5%?

	Edad		
Ve	16–34	35–54	55 y más
Poca violencia	8	12	21
Mucha violencia	18	15	7

14.19 Los resultados de un estudio[8] sugieren que el electrocardiograma (ECG) inicial de una víctima que se sospecha sufre un ataque al corazón se pueden usar para predecir complicaciones de naturaleza aguda en el hospital. El estudio incluyó 469 pacientes con sospecha de infarto al miocardio (ataque al corazón). Cada uno de los pacientes fue clasificado de acuerdo con si su ECG inicial era positivo o negativo y si la persona sufría complicaciones que ponían en riesgo su vida después en el hospital. Los resultados se resumen en la siguiente tabla.

	Complicaciones que ponen en riesgo la vida del internado en hospital		
ECG	No	Sí	Total
Negativo	166	1	167
Positivo	260	42	302
Total	426	43	469

a ¿Hay suficiente evidencia para indicar que si el paciente de un ataque al corazón sufre o no complicaciones depende del resultado del ECG inicial? Pruebe usando $\alpha = .05$.

b Establezca límites para el nivel de significancia alcanzado.

14.20 Consulte el Ejercicio 14.10. Pruebe la hipótesis, con un nivel de significancia de 5%, que los defectos del tipo A se presentan independientemente de los defectos tipo B.

14.21 Un uso interesante y práctico de la prueba χ^2 ocurre al probar la segregación de especies de plantas o animales. Suponga que dos especies de plantas, A y B, están creciendo en un lote de prueba. Para evaluar si las especies tienden a segregarse, un investigador muestrea al azar n plantas del lote; se registran la especie de cada planta muestreada *y* la especie de su vecina más cercana. Los datos se anotan luego en una tabla, como se ilustra aquí.

8. *Fuente*: J. E. Brush y otros, "Use of the Initial Electrocardiogram to Predict In-Hospital Complications of Acute Myocardial Infarction," *New England Journal of Medicine* (mayo de 1985).

	Vecina más cercana	
Planta muestreada	A	B
A	a	b
B	c	d
		n

Si a y d son grandes con respecto a b y c, estaríamos inclinados a decir que las especies tienden a segregarse. (Casi todas las vecinas de las A son del tipo A, y casi todas las vecinas de las B son del tipo B.) Si b y c son grandes en comparación con a y d, diríamos que las especies tienden a estar demasiado revueltas. En cualquiera de estos casos (segregación o demasiado revueltas), una prueba χ^2 debe dar un valor grande y la hipótesis de mezcla aleatoria se rechazaría. Para cada uno de los casos siguientes, pruebe la hipótesis de mezcla aleatoria (o bien, de manera equivalente, la hipótesis que afirma que la especie de una planta es independiente de la especie de su vecina más cercana). Use $\alpha = .05$ en cada caso.

a $a = 20, b = 4, c = 8, d = 18$.
b $a = 4, b = 20, c = 18, d = 8$.
c $a = 20, b = 4, c = 18, d = 8$.

14.5 Tablas $r \times c$ con totales fijos de renglón o columna

En la sección anterior, describimos el análisis de una tabla de contingencia $r \times c$ mediante el uso de ejemplos que, para todos los fines prácticos, se ajusta al experimento multinomial descrito en la Sección 14.1. Aunque los métodos para recolectar datos en muchos estudios pueden satisfacer los requisitos de un experimento multinomial, otros métodos no los satisfacen. Por ejemplo, podríamos no querer muestrear aleatoriamente la población descrita en el Ejemplo 14.3 porque podríamos hallar que, debido a la casualidad, faltaría una categoría completa. Las personas que no han recibido inyecciones de vacuna contra la gripe podrían no aparecer en la muestra. Podríamos decidir de antemano entrevistar a un número especificado de personas de cada categoría de columna, con lo cual se fijan anticipadamente los totales de columna. Entonces tendríamos tres experimentos binomiales separados e independientes, correspondientes a “no vacuna”, “una vacuna” y “dos vacunas”, con probabilidades respectivas de p_1, p_2 y p_3 de que una persona contraiga la gripe. En este caso, estamos interesados en probar la hipótesis nula

$$H_0 : p_1 = p_2 = p_3.$$

(En realidad estamos probando la equivalencia de tres distribuciones binomiales.) De acuerdo con esta hipótesis, los MLE de las frecuencias esperadas por celda son los mismos que en la Sección 14.4, es decir,

$$\widehat{E(n_{ij})} = \frac{r_i c_j}{n}.$$

¿Cuántos grados de libertad están asociados con la distribución de aproximación χ^2? Hay $r\,c$ probabilidades en total. Como los totales de columna son fijos, la suma de las probabilidades *en cada columna* debe ser igual a uno. Esto es,

$$p_{1j} + p_{2j} + \cdots\ p_{rj} = 1, \qquad \text{para cada } j = 1, 2, \ldots\ c,$$

y hay c restricciones lineales en las p_{ij}, que resultan en una pérdida de c grados de libertad. Finalmente, es necesario estimar $r - 1$ probabilidades de renglón (las probabilidades estimadas de renglón deben sumar 1), disminuyendo los grados de libertad en una $r - 1$ adicional. Así, el número de grados de libertad asociados con X^2 calculado para una tabla $r \times c$ con totales fijos de columna es gl $= rc - c - (r - 1) = (r - 1)(c - 1)$.

Para ilustrar, suponga que deseamos probar una hipótesis respecto a la equivalencia de cuatro poblaciones binomiales, como se indica en el siguiente ejemplo.

EJEMPLO **14.4** Una encuesta de las opiniones de los votantes se realizó en cuatro distritos políticos urbanos para comparar la fracción de votantes que están a favor del candidato A. Muestras aleatorias de 200 votantes fueron entrevistados en cada uno de los cuatro distritos, con los resultados que se muestran en la Tabla 14.5. ¿Los datos presentan suficiente evidencia para indicar que las fracciones de votantes a favor del candidato A difieren en los cuatro distritos?

Solución Observará que la mecánica para probar las hipótesis respecto a la equivalencia de los parámetros de las cuatro poblaciones binomiales, que corresponden a los cuatro distritos, es idéntica a la mecánica asociada con probar la hipótesis de independencia de las clasificaciones de renglón y columna. Si denotamos la fracción de votantes a favor de A como p y planteamos la hipótesis de que p es igual para los cuatro distritos, implicamos que las probabilidades del primer renglón son todas iguales a p y que las probabilidades del segundo renglón son todas iguales a $1 - p$. El MLE (que combina los resultados de las cuatro muestras) para el valor común de p es $\hat{p} = 236/800 = r_1/n$. El número esperado de personas que están a favor de A en el distrito 1 es $E(n_{11}) = 200p$, que es estimado por el valor

$$\widehat{E(n_{11})} = 200\hat{p} = 200\left(\frac{236}{800}\right) = \frac{(c_1 r_1)}{n}.$$

Observe que aun cuando estamos analizado un experimento muy diferente al considerado en la Sección 14.4, las frecuencias medias estimadas por celda se calculan del mismo modo que en la Sección 14.4. Las otras frecuencias por celda esperadas y estimadas, calculadas con el uso de los totales de renglón y columna, aparecen en la Tabla 14.5. Vemos que

$$\begin{aligned} X^2 &= \sum_{j=1}^{4}\sum_{i=1}^{2} \frac{\left[n_{ij} - \widehat{E(n_{ij})}\right]^2}{\widehat{E(n_{ij})}} \\ &= \frac{(76-59)^2}{59} + \frac{(124-141)^2}{141} + \cdots + \frac{(152-141)^2}{141} = 10.72. \end{aligned}$$

Tabla 14.5 Tabulación de datos para el Ejemplo 14.4

	Distrito				
Opinión	1	2	3	4	Total
A favor de A	76 (59)	53 (59)	59 (59)	48 (59)	236
No a favor de A	124 (141)	147 (141)	141 (141)	152 (141)	564
Total	200	200	200	200	800

El valor crítico de χ^2 para $\alpha = .05$ y $(r - 1)(c - 1) = (1)(3) = 3$ grados de libertad es 7.815. Como X^2 excede este valor crítico, rechazamos la hipótesis nula y concluimos que la fracción de votantes a favor del candidato A no es igual para todos los distritos. El valor p asociado está dado por $P(\chi^2 > 10.72)$ cuando χ^2 tiene 3 grados de libertad. Así, $.01 \leq$ valor $p \leq .025$. La aplicación χ^2 da $P(\chi^2 > 10.72) = .01334$. ■

Este ejemplo fue resuelto en el Ejercicio 10.106 por el método de razón de probabilidades. Observe que las conclusiones son las mismas.

La prueba realizada en el Ejemplo 14.4 es una prueba de la igualdad de cuatro proporciones binomiales con base en muestras independientes a partir de cada una de las poblaciones correspondientes. Es frecuente que esta prueba se denomine *prueba de homogeneidad* de las poblaciones binomiales. Si hay más de dos categorías de renglón y los totales de columna son fijos, la prueba χ^2 es una prueba de la equivalencia de las proporciones en c poblaciones multinomiales.

Ejercicios

14.22 Un estudio para determinar la efectividad de un medicamento (suero) para el tratamiento de la artritis resultó en la comparación de dos grupos, cada uno de ellos formado por 200 pacientes artríticos. Un grupo fue inoculado con el suero mientras que el otro recibió un placebo (inoculación que parece contener suero pero que en realidad no es activo). Después de un tiempo, a cada persona del estudio se le preguntó si su condición artrítica había mejorado; se observaron los siguientes resultados. ¿Estos datos presentan suficiente evidencia para indicar que diferían las proporciones de personas artríticas que dijeron que habían mejorado, dependiendo de si habían recibido el suero?

Condición	Tratado	No tratado
Mejor	117	74
No mejor	83	126

a Pruebe usando el estadístico X^2. Use $\alpha = .05$.

b Pruebe usando la prueba Z de la Sección 10.3 y $\alpha = .05$. Compare su resultado con el del inciso a.

c Establezca límites para el nivel de significancia alcanzado con la prueba del inciso a.

14.23 La prueba χ^2 usada en el Ejercicio 14.22 es equivalente a la prueba Z de dos colas de la Sección 10.3, siempre que α sea la misma para las dos pruebas. Demuestre algebraicamente que el estadístico X^2 de la prueba χ^2 es el cuadrado del estadístico de la prueba Z para la prueba correspondiente.

14.24 ¿En qué forma los estadounidenses de la "generación sándwich" equilibran las demandas de atención para familiares más viejos y más jóvenes? La siguiente tabla contiene los resultados de una encuesta por teléfono de estadounidenses entre 45 y 55 años llevada a cabo por *The New York Times*.[9] De cada una de cuatro subpoblaciones se encuestaron 200 personas y se les preguntó si daban apoyo financiero a sus padres.

9. *Fuente*: adaptado de Tamar Lewin, "Report Looks at a Generation, and Caring for Young and Old," *New York Times* en línea, 11 de julio de 2001.

	Subpoblación			
Apoyo	Estadounidenses blancos	Afroamericanos	Hispano-americanos	Asiáticoamericanos
Sí	40	56	68	84
No	160	144	132	116

a Use la prueba χ^2 para determinar si las proporciones de personas que dan apoyo financiero a sus padres difiere para las cuatro subpoblaciones. Use $\alpha = .05$.

b Como las muestras son independientes, los intervalos de confianza para comparar las proporciones de cada subpoblación que apoya financieramente a sus padres se pueden obtener usando el método presentado en la Sección 8.6.

i Establezca un intervalo de confianza de 95% para la diferencia en las proporciones que dan apoyo a sus padres para estadounidenses blancos y asiáticos.

ii Use el método Bonferroni presentado en la Sección 13.12 para dar seis intervalos simultáneos de confianza con el fin de comparar las proporciones que dan apoyo a sus padres para todos los pares de subpoblaciones. El objetivo es dar intervalos con coeficiente simultáneo de confianza de al menos .95.

iii Con base en su respuesta al inciso ii, ¿cuáles son las subpoblaciones que difieren de las otras con respecto a la proporción de las que dan apoyo financiero a sus padres?

14.25 ¿La educación hace realmente la diferencia respecto a cuánto dinero se gana? Unos investigadores seleccionaron aleatoriamente a 100 personas de cada una de tres categorías de ingresos: "marginalmente ricos", "cómodamente ricos" y "superricos", y registraron sus niveles de educación. Los datos se resumen en la siguiente tabla.[10]

Nivel de educación máximo	Marginalmente rico	Cómodamente rico	Superrico
Sin universidad	32	20	23
Algunos estudios universitarios	13	16	1
Estudiante sin graduar	43	51	60
Estudios de posgrado	12	13	16
Total	100	100	100

a Describa las poblaciones multinomiales independientes cuyas proporciones están comparadas en el análisis χ^2.

b ¿Los datos indican que las proporciones de los diversos niveles de educación difieren para las tres categorías de ingresos? Pruebe en el nivel $\alpha = .01$.

c Construya un intervalo de confianza de 95% para la diferencia en proporciones con al menos un estudiante sin graduar para personas que son marginalmente ricos y superricos. Interprete el intervalo.

14.26 Un fabricante de botones deseaba determinar si la fracción de botones defectuosos producidos por tres máquinas variaba de una máquina a otra. Se seleccionaron muestras de 400 botones de cada una de las tres máquinas y se contó el número de defectuosos por cada muestra. Los resultados se observan en la siguiente tabla. ¿Estos datos presentan suficiente evidencia para indicar que la fracción de botones defectuosos variaba de una máquina a otra?

10. *Fuente*: adaptado de Rebecca Piirto Heath, "Life on Easy Street," *American Demographics,* abril de 1997, p. 33.

Máquina número	Número de botones defectuosos
1	16
2	24
3	9

a Pruebe, usando $\alpha = .05$, con una prueba χ^2.

***b** Pruebe, usando $\alpha = .05$, con una prueba de razón de probabilidades. [*Sugerencia*: consulte el Ejercicio 10.106.][11]

14.27 H. W. Menard[12] realizó una investigación respecto a nódulos de manganeso, aglomeración rica en minerales que se encuentra en abundancia en el lecho marino a gran profundidad. En una parte de su informe, Menard proporcionó datos relacionados con la edad magnética de la corteza terrestre y "la probabilidad de hallar nódulos de manganeso". Los datos mostrados en la siguiente tabla representan el número de muestras del núcleo de la tierra, así como el porcentaje de las que contenían nódulos de manganeso para cada una de varias edades magnéticas de la corteza terrestre. ¿Los datos arrojan suficiente evidencia para indicar que la probabilidad de hallar nódulos de manganeso en la corteza difiere dependiendo de la clasificación de la edad magnética? Pruebe con $\alpha = .05$.

Edad	Número de muestras	Porcentaje con nódulos
Mioceno, reciente	389	5.9
Oligoceno	140	17.9
Eoceno	214	16.4
Paleoceno	84	21.4
Cretácico tardío	247	21.1
Cretácico temprano y medio	1120	14.2
Jurásico	99	11.0

14.28 Tradicionalmente, los sindicatos de trabajadores en Estados Unidos han estado satisfechos con dejar la administración de empresas a gerentes y ejecutivos corporativos. En Europa, la participación de los trabajadores en la toma de decisiones administrativas es una idea aceptada que se hace cada vez más popular. Para estudiar el efecto de la participación de los trabajadores, se entrevistó a 100 trabajadores en cada una de dos plantas manufactureras alemanas. Una planta tenía participación activa de trabajadores en la toma de decisiones gerenciales, no así la otra. A cada trabajador seleccionado se le preguntó si aprobaba las decisiones gerenciales tomadas dentro de la planta. Los resultados aparecen a continuación.

	Participación	Sin participación
Generalmente aprueban	73	51
No aprueban	27	49

a ¿Los datos indican una diferencia en las proporciones de trabajadores de las dos plantas que generalmente aprueban las decisiones gerenciales? Pruebe con un nivel de significancia de .05 usando la prueba χ^2.

b Construya un límite inferior de confianza de 95% para la diferencia en la proporción de trabajadores que aprueban decisiones gerenciales en las plantas con y sin su participación.

11. Los ejercicios precedidos por un asterisco son opcionales.

12. *Fuente*: H. W. Menard, "Time, Chance and the Origin of Manganese Nodules," *American Scientist*, septiembre-octubre de 1976. ©1976 Scientific Research Company of North America. Todos los derechos reservados.

¿El límite de confianza resultante indica que una mayor proporción de trabajadores aprueba decisiones gerenciales en la planta con su participación activa? ¿Por qué?

c ¿La conclusión a la que se llegó en el inciso b podría haber resultado de la prueba χ^2 realizada en el inciso a? ¿Por qué?

14.29 Se realizó una inspección para estudiar la relación entre enfermedades pulmonares y la contaminación del aire. Se escogieron cuatro lugares para la inspección, dos ciudades que con frecuencia sufren de contaminación por el esmog y dos ciudades no urbanas en estados con bajos índices de contaminación del aire. Se incluyeron sólo adultos residentes permanentes de los lugares del estudio. Muestras aleatorias de 400 adultos residentes permanentes de cada uno de los lugares dieron los resultados que aparecen en la siguiente tabla.

Área	Número de enfermedades pulmonares
Ciudad A	34
Ciudad B	42
Zona no urbana 1	21
Zona no urbana 2	18

a ¿Los datos aportan suficiente evidencia para indicar una diferencia en las proporciones con enfermedades pulmonares para los cuatro lugares?

b ¿Los fumadores debieron excluirse de las muestras? ¿Cómo afectaría esto a las inferencias obtenidas de los datos?

14.30 Consulte el Ejercicio 14.29. Calcule la diferencia en las fracciones de residentes permanentes adultos con enfermedades pulmonares para las ciudades A y B. Use un intervalo de confianza de 95%.

14.31 Se realizó un estudio para investigar el interés de adultos de edad mediana en los programas de acondicionamiento físico en Rhode Island, Colorado, California y Florida. El objetivo de la investigación era determinar si la participación de adultos en programas de acondicionamiento físico varía de una región de Estados Unidos a otra. Muestras aleatorias de personas fueron entrevistadas en cada estado y se registraron los datos obtenidos en la siguiente tabla. ¿Los datos indican diferencias entre los porcentajes de participación de adultos en estos programas de un estado a otro? ¿Qué se concluiría con $\alpha = .01$?

Participación	Rhode Island	Colorado	California	Florida
Sí	46	63	108	121
No	149	178	192	179

14.6 Otras aplicaciones

Las aplicaciones de la prueba χ^2 para el análisis de datos categóricos descritos en las Secciones 14.3–14.5 representan sólo algunos de los interesantes problemas de clasificación que pueden calcularse mediante el experimento multinomial y para el cual nuestro método de análisis es apropiado. Por lo general estas aplicaciones son complicadas en mayor o menor grado porque los valores numéricos de las probabilidades por celda no están especificados y, por tanto, requieren la estimación de uno o más parámetros poblacionales. Entonces, al igual que en las Secciones 14.4 y 14.5, podemos calcular las probabilidades por celda. Aun cuando omitimos la mecánica de las pruebas estadísticas, varias aplicaciones adicionales de la prueba χ^2 merecen citarse como tema de interés.

Por ejemplo, suponga que deseamos probar una hipótesis que afirma que una población posee una distribución normal de probabilidad. Las celdas de un histograma de frecuencia muestral corresponderían a las k celdas del experimento multinomial y las frecuencias observadas por celda serían el número de mediciones que caen en cada celda del histograma. Dada la hipotética distribución de probabilidad normal para la población, podríamos usar las áreas bajo la curva normal para calcular las probabilidades teóricas por celda y en consecuencia las frecuencias esperadas por celda. Los MLE deben emplearse cuando μ y σ no estén especificados para la población normal y estos parámetros deben calcularse para estimar probabilidades por celda.

La construcción de una tabla de dos direcciones para investigar la dependencia entre dos clasificaciones se puede ampliar a tres o más clasificaciones. Por ejemplo, si deseamos probar la independencia mutua de tres clasificaciones, emplearíamos una "tabla" de tres dimensiones. El razonamiento y metodología asociados con el análisis de tablas de dos y tres entradas son idénticos aun cuando el análisis de la tabla de tres entradas sea un poco más complejo.

Una tercera e interesante aplicación de nuestra metodología sería su uso en la investigación de la rapidez de cambio de una población multinomial (o binomial) como función del tiempo. Por ejemplo, podríamos estudiar la capacidad para resolver el problema que tenga una persona (o cualquier animal) sometida a un programa educacional y examinada en el tiempo. Si, por ejemplo, la persona es examinada a intervalos prescritos y el examen es del tipo sí o no, dando varias respuestas correctas y que seguirían una distribución binomial de probabilidad, estaríamos interesados en el comportamiento de la probabilidad de una respuesta correcta p como función del tiempo. Si el número de respuestas correctas se registró para c periodos, los datos caerían en una tabla de $2 \times c$ semejante a la del Ejemplo 14.4 (Sección 14.5). Entonces estaríamos interesados en probar la hipótesis de que p es igual a una constante, es decir, que no ha ocurrido aprendizaje y entonces plantearíamos hipótesis más interesantes para determinar si los datos presentan suficiente evidencia para indicar un cambio gradual (lineal, por ejemplo) en el tiempo, en oposición a un cambio abrupto en algún punto en el tiempo. Los procedimientos que hemos descrito podrían ampliarse a decisiones que comprendan más de dos alternativas.

Usted observará que nuestro ejemplo de cambio en el tiempo es común en las finanzas, la industria y en muchos otros campos de actividad, incluyendo las ciencias sociales. Por ejemplo, podríamos desear estudiar el porcentaje de aceptación de los consumidores de un producto nuevo para varios tipos de campañas publicitarias como función del tiempo que la campaña haya estado en vigor. O podríamos estudiar la tendencia de la fracción de un lote de piezas defectuosas en un proceso de manufactura como función del tiempo. Estos dos ejemplos, así como muchos otros, requieren un estudio del comportamiento de un proceso binomial (o multinomial) como función del tiempo.

Los ejemplos que acabamos de describir tienen la intención de sugerir la aplicación relativamente amplia del análisis χ^2 de datos categóricos, un hecho que debe ser recordado por el experimentador que se ocupe de este tipo de datos. Es frecuente que la prueba estadística que utilice X^2 como estadístico de prueba sea denominada *prueba de bondad de ajuste.* Su aplicación para algunos de los ejemplos requiere cuidado en la determinación de las estimaciones apropiadas y el número de grados de libertad para X^2, que en algunos de estos problemas puede resultar muy complicado.

14.7 Resumen y conclusiones

El material de este capítulo se ha referido a pruebas de hipótesis relacionadas con probabilidades por celda asociadas con experimentos multinomiales (Secciones 14.2 y 14.3) o con varios experimentos multinomiales independientes (Sección 14.5). Cuando el número de observaciones n es grande, se puede demostrar que el estadístico de prueba X^2 posee, aproximadamente, una distribución de probabilidad χ^2 en muestreo repetido, el número de grados de libertad dependerá de la aplicación particular. En general, suponemos que n es grande y que la frecuencia mínima esperada por celda es igual o mayor que cinco.

Unas palabras de advertencia respecto al uso del estadístico X^2 como método para analizar datos categóricos. La determinación del número correcto de grados de libertad asociados con el estadístico X^2 es crítica para localizar la región de rechazo. Si el número se especifica incorrectamente, podrían resultar conclusiones erróneas. Observe también que el no rechazo de la hipótesis nula no implica que deba ser aceptada. Tendríamos dificultad para expresar una hipótesis alternativa significativa para muchas aplicaciones prácticas y, por tanto, nos faltaría conocimiento de la probabilidad de cometer un error tipo II. Por ejemplo, expresamos la hipótesis de que dos clasificaciones de una tabla de contingencia son independientes. Una alternativa específica debe determinar una medida de dependencia que puede o no poseer significancia práctica para el experimentador. Por último, si faltan parámetros y deben estimarse las frecuencias esperadas por celda, los parámetros faltantes deben calcularse por el método de máxima probabilidad para que la prueba sea válida. En otras palabras, el empleo de la prueba χ^2 para otras aplicaciones que no sean las indicadas en las Secciones 14.3–14.5 requerirá de una experiencia que está fuera del propósito de esta presentación introductoria del tema.

Bibliografía y lecturas adicionales

Agresti, Alan. 2002. *Categorical Data Analysis,* 2d ed. New York: Wiley-Interscience.

Agresti, Alan. 2007. *An Introduction to Categorical Data Analysis,* 2d ed. New York: Wiley-Interscience.

Cochran, W. G. 1952. "The χ^2 Test of Goodness of Fit", *Annals of Mathematical Statistics* 23: 315–345.

Conover, W. J. 1999. *Practical Nonparametric Statistics,* 3d ed. New York: Wiley.

Daniel, W. W. 1990. *Applied Nonparametric Statistics,* 2d ed. Boston: PWS-Kent.

Fienberg, Stephen E. 1980. The *Analysis of Cross-Classified Categorical Data,* 2d ed. Cambridge, Mass.: MIT Press.

Kendall, M. G., A. Stuart, J. K. Ord, and S. Arnold. 1999. *Kendall's Advanced Theory of Statistics: Volume 2A—Classical Inference and the Linear Model,* 6th ed. London: Arnold.

Ejercicios complementarios

14.32 Mencione las características de un experimento multinomial.

14.33 Se llevó a cabo un estudio para determinar las actitudes de estudiantes, profesores y empleados administrativos acerca de una nueva política de estacionamientos en una universidad. La distribución de quienes están a favor o se oponen a esa política se muestran en la siguiente tabla. ¿Los datos dan suficiente evidencia para indicar que las actitudes respecto a la política de estacionamiento son independientes de la condición de estudiante, profesor o empleado de administración?

Opinión	Estudiante	Profesor	Administración
A favor	252	107	43
Se opone	139	81	40

14.34 ¿Cómo se califica usted como automovilista? De acuerdo con un estudio llevado a cabo por el Field Institute,[13] la mayoría de los californianos piensan que son buenos conductores pero que tienen poco respeto por la capacidad de conducir de otros. Los datos de las tablas siguientes muestran la distribución de opiniones, según el género, para dos preguntas diferentes. Los datos de la primera tabla proporcionan los resultados obtenidos cuando los conductores se califican a sí mismos; la segunda tabla muestra los resultados obtenidos cuando los conductores calificaron a otros. Aun cuando no se cita en la fuente, suponemos que hubo 100 hombres y 100 mujeres en cada uno de los grupos estudiados.

Opinión de sí mismo como conductor			
Género	Excelente	Bueno	Regular
Hombre	43	48	9
Mujer	44	53	3

Opinión de otros conductores				
Género	Excelente	Bueno	Regular	Malo
Hombre	4	42	41	13
Mujer	3	48	35	14

a Consulte la tabla en la que los conductores se califican a sí mismos. ¿Hay suficiente evidencia para indicar que hay una diferencia en las proporciones de las tres categorías de calificación para automovilistas hombres y mujeres? Establezca límites para el valor p asociado con la prueba.

b Consulte la tabla en la que los conductores califican a otros. ¿Hay suficiente evidencia para indicar que hay una diferencia en las proporciones de las cuatro categorías de calificación cuando califican a automovilistas hombres y mujeres? Establezca límites para el valor p asociado con la prueba.

c ¿Violó usted alguna de las suposiciones en sus análisis en los incisos a y b? ¿Qué efecto podrían tener estas violaciones sobre la validez de las conclusiones?

13. *Fuente*: Dan Smith, "Motorists Have Little Respect for Others' Skills," *Press-Enterprise* (Riverside, Calif.) 15 de marzo de 1991.

14.35 ¿La probabilidad de "pescar" un resfriado está influenciada por el número de contactos sociales que tiene una persona? Un estudio de Sheldon Cohen, profesor de psicología de la Universidad Carnegie Melon, parece demostrar que cuantas más relaciones sociales tenga una persona, *es menos susceptible* a resfriados. Un grupo de 276 hombres y mujeres sanos se agruparon de acuerdo con su número de relaciones (por ejemplo como padre, amigo, miembro de la iglesia, vecino). A continuación se expusieron a un virus causante del resfriado. En la tabla siguiente se da una adaptación de los resultados.[14]

	Número de relaciones		
	3 o menos	4 o 5	6 o más
Resfriado	49	43	34
No resfriado	31	57	62
Total	80	100	96

a ¿Los datos presentan suficiente evidencia para indicar que la susceptibilidad a resfriados es afectada por el número de relaciones que tienen las personas? Pruebe con un nivel de significancia de 5%.

b Establezca límites para el valor p.

14.36 Las lesiones en las rodillas pueden ser un problema importante para los atletas en muchos deportes de contacto. No obstante, aquellos atletas que juegan ciertas posiciones son más propensos a lesiones en las rodillas que otros jugadores. La prevalencia y las formas de lesiones en rodillas entre jugadoras colegiales de rugby fueron investigadas por medio de un sencillo cuestionario, al cual respondieron 42 clubes de rugby.[15] Un total de 76 lesiones en las rodillas se clasificó por tipo y posición (delantero o defensa) del jugador lesionado.

Posición	Desgarre de meniscos	Desgarre MCL	Desgarre ACL	Otras
Delantero	13	14	7	4
Defensa	12	9	14	3

a ¿Los datos dan suficiente evidencia para indicar dependencia entre la posición jugada y el tipo de lesiones en las rodillas? Pruebe usando $\alpha = .05$.

b Dé límites para el valor p asociado con el valor para X^2 obtenido en el inciso a.

c **Ejercicio Applet** Use la aplicación *Chi-Square Probability and Quantiles* para determinar el valor p asociado con el valor de X^2 obtenido en el inciso a.

14.37 Con frecuencia no queda claro si todas las propiedades de un experimento binomial se satisfacen en realidad en una aplicación determinada; una prueba de bondad de ajuste es deseable para estos casos. Suponga que un experimento formado por cuatro ensayos se repite 100 veces. El número de repeticiones en las que se obtuvo un número dado de éxitos se registra en la siguiente tabla. Estime p (suponiendo que el experimento haya sido binomial), obtenga estimaciones de las frecuencias esperadas por celda y

14. *Fuente*: adaptado de David L. Wheeler, "More Social Roles Means Fewer Colds," *Chronicle of Higher Education* 43(44) (1997): A13.

15. *Fuente*: Andrew S. Levy, M. J. Wetzler, M. Lewars, and W. Laughlin, "Knee Injuries in Women Collegiate Rugby Players," *American Journal of Sports Medicine* 25(3) (1997): 360.

pruebe la bondad de ajuste. Para determinar el número apropiado de grados de libertad para X^2, observe que p tuvo que ser estimada.

Posibles resultados (número de éxitos)	Número de veces que se obtuvo
0	11
1	17
2	42
3	21
4	9

14.38 El conteo del número de artículos por conjunto (colonia o grupo) debe ser necesariamente mayor o igual a 1. Así, la distribución de Poisson generalmente no se ajusta a estas clases de conteos. Para modelar conteos sobre fenómenos como son el número de bacterias por colonia, el número de personas por familia, así como el número de animales por camada, es frecuente que la distribución de la *serie logarítmica* resulte útil. Esta distribución discreta tiene función de probabilidad dada por

$$p(y \mid \theta) = -\frac{1}{\ln(1-\theta)} \frac{\theta^y}{y}, \qquad y = 1, 2, 3, \ldots, \; 0 < \theta < 1,$$

donde θ es un parámetro desconocido.

a Demuestre que el MLE $\hat{\theta}$ de θ satisface la ecuación

$$\overline{Y} = \frac{\hat{\theta}}{-(1-\hat{\theta})\ln(1-\hat{\theta})}, \qquad \text{donde} \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

b Los datos de la siguiente tabla representan frecuencias observadas sobre el número de bacterias por colonia, para cierto tipo de bacterias del suelo.[16]

Bacterias por colonia	1	2	3	4	5	6	7+
Número de colonias observado	359	146	57	41	26	17	29

Pruebe la hipótesis de que estos datos se ajustan a una distribución de serie logarítmica. Use $\alpha = .05$. (Observe que el valor $\overline{y}$ debe ser aproximado porque no tenemos información exacta en recuentos mayores que seis.)

14.39 Consulte la tabla de contingencia $r \times c$ de la Sección 14.4. Demuestre que el MLE de la probabilidad p_i para el renglón i es $\hat{p}_i = r_i/n$, para $i = 1, 2, \ldots, r$.

***14.40** Un modelo genético expresa que las proporciones de descendientes en tres clases debe ser p^2, $2p(1-p)$ y $(1-p)^2$ para un parámetro p, $0 \le p \le 1$. Un experimento arrojó frecuencias de 30, 40 y 30 para las clases respectivas.

a ¿El modelo se ajusta a los datos? (Use máxima probabilidad para estimar p.)

b Suponga que la hipótesis expresa que el modelo se cumple con $p = .5$. ¿Los datos contradicen esta hipótesis?

***14.41** De acuerdo con el modelo genético para la relación entre sexo y ceguera a los colores, las cuatro categorías, macho y normal, hembra y normal, macho y ciego al color, hembra y ciego al color, deben tener

16. *Fuente*: C. A. Bliss and R. A. Fisher, "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics* 9 (1953): 176-200. Biometrics Society. Todos los derechos reservados.

probabilidades dadas por $p/2$, $(p^2/2) + pq$, $q/2$ y $q^2/2$ respectivamente, donde $q = 1 - p$. Una muestra de 2000 personas reveló 880, 1032, 80 y 8 en las categorías respectivas. ¿Estos datos concuerdan con el modelo? Use $\alpha = .05$. (Use máxima probabilidad para estimar p.)

***14.42** Suponga que $(Y_1, Y_2, \ldots, Y_k)$ tiene una distribución multinomial con parámetros $n, p_1, p_2, \ldots, p_k$ y $(X_1, X_2, \ldots, X_k)$ tiene una distribución multinomial con parámetros $m, p_1^*, p_2^*, \ldots, p_k^*$. Construya una prueba de la hipótesis nula de que las dos distribuciones multinomiales son idénticas; esto es, pruebe $H_0 : p_1 = p_1^*, p_2 = p_2^*, \ldots, p_k = p_k^*$.

***14.43** En un experimento para evaluar un insecticida se esperaba que la probabilidad de que sobrevivan insectos fuera lineal con respecto a la dosis D en la región de experimentación; esto es, $p = 1 + \beta D$. Se realizó un experimento usando cuatro niveles de dosis, 1, 2, 3 y 4, y 1000 insectos en cada grupo. Los datos resultantes se muestran en la tabla siguiente. ¿Estos datos contradicen la hipótesis de que $p = 1 + \beta D$? [*Sugerencia*: escriba las probabilidades por celda en términos de β y encuentre el MLE de β.]

Dosis	Número de sobrevivientes
1	820
2	650
3	310
4	50

CAPÍTULO 15

Estadística no paramétrica

15.1 Introducción

Algunos experimentos dan como resultado medidas de respuesta que desafían una cuantificación exacta. Por ejemplo, suponga que se le pide a un juez evaluar y clasificar la capacidad de instrucción de cuatro profesores, o las características comestibles y de sabor de cinco marcas de hojuelas de maíz. Debido a que es claramente imposible dar una medida exacta de la competencia de un profesor o el sabor de un alimento, las mediciones de respuesta son de un carácter completamente diferente a las presentadas en los capítulos anteriores. En ejemplos como éstos, los experimentos generan mediciones de respuesta que se pueden ordenar (clasificar), pero es imposible hacer enunciados tales como “el profesor A es el doble de bueno que el profesor B”, Aunque experimentos de este tipo se presentan en casi todos los campos de actividad, son particularmente evidentes en la investigación en ciencias sociales y en estudios de preferencias de consumidores. Los métodos estadísticos no paramétricos son útiles para analizar este tipo de datos.

Los procedimientos estadísticos no paramétricos se aplican no sólo a observaciones que son difíciles de cuantificar, sino que también son particularmente útiles para hacer inferencias en situaciones en las que existe una seria duda acerca de las suposiciones que son la base de la metodología estándar. Por ejemplo, la prueba t para comparar un par de medias basadas en muestras independientes, Sección 10.8, está basada en la suposición de que ambas poblaciones están distribuidas normalmente con varianzas iguales. El experimentador nunca sabrá si estas suposiciones se cumplen en una situación práctica, pero con frecuencia estará razonablemente seguro de que las desviaciones de las suposiciones serán lo suficientemente pequeñas para que las propiedades del procedimiento estadístico no se alteren. Esto es, α y β serán más o menos lo que el experimentador piensa que son. Por otra parte, no es raro que el experimentador dude de la validez de una suposición y se pregunte si está usando un procedimiento estadístico válido. A veces esta dificultad se puede salvar si se usa una prueba de estadístico no paramétrico, con lo cual se evita usar un procedimiento estadístico que sea apropiado sólo para un conjunto muy incierto de suposiciones.

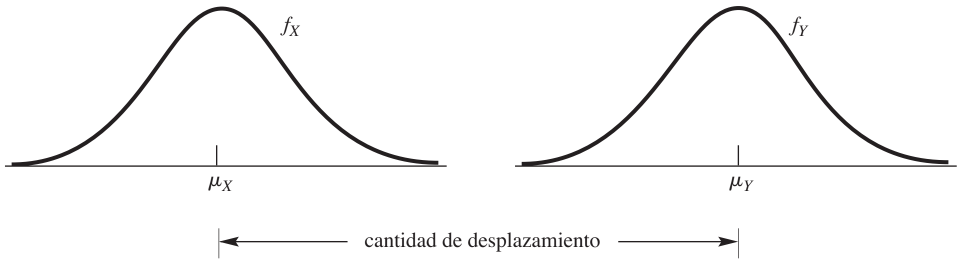
El término *estadístico no paramétrico* no tiene definición estándar que haya sido acordada por todos los expertos en estadística, pero casi todos concuerdan en que los métodos estadísticos no paramétricos funcionan bien con suposiciones bastante generales acerca de la naturaleza de cualesquiera distribuciones de probabilidad o parámetros que intervienen en un problema inferencial. Como tesis de trabajo, definiremos *métodos paramétricos* como aquellos que se aplican a problemas donde la(s) distribución(es) de la(s) cual(es) se toman las muestras están especificadas excepto para los valores de un número finito de parámetros. Los métodos no paramétricos se aplican en todos los otros casos. Por ejemplo, la prueba t de una muestra desarrollada en el Capítulo 10 se aplica cuando la población está distribuida normalmente con media y varianza desconocidas. Debido a que la distribución de la cual se toma la muestra está especificada excepto para los valores de dos parámetros, μ y σ^2, la prueba t es un procedimiento paramétrico. En forma alternativa, suponga que muestras independientes se toman de dos poblaciones y que deseamos probar la hipótesis de que dos distribuciones poblacionales son idénticas pero de forma no especificada. En este caso, la distribución es no especificada y la hipótesis debe ser probada con el uso de métodos no paramétricos.

El empleo válido de algunos de los métodos paramétricos presentados en capítulos anteriores exige que se satisfagan al menos aproximadamente ciertas suposiciones de distribución. Incluso si se satisfacen todas las suposiciones, la investigación ha demostrado que las pruebas estadísticas no paramétricas son casi tan capaces de detectar diferencias entre poblaciones como los métodos paramétricos aplicables. Pueden ser, y con frecuencia lo son, más potentes para detectar diferencias poblacionales cuando las suposiciones no se satisfacen. Por esta razón, muchos expertos en estadística están a favor del uso de procedimientos estadísticos no paramétricos en vez de sus equivalentes paramétricos.

15.2 Modelo general de desplazamiento (o cambio) de dos muestras

Es común que un experimentador tome observaciones de dos poblaciones con el fin de probar si tienen la misma distribución. Por ejemplo, si muestras aleatorias independientes $X_1, X_2, \dots, X_{n_1}$ y $Y_1, Y_2, \dots, Y_{n_2}$ se toman de poblaciones normales con iguales varianzas y medias respectivas μ_X y μ_Y, el experimentador puede desear probar $H_0: \mu_X - \mu_Y = 0$ contra $H_a: \mu_X - \mu_Y < 0$. En este caso, si H_0 es verdadera, ambas poblaciones están distribuidas normalmente con

FIGURA **15.1** Dos distribuciones normales con varianzas iguales pero medias desiguales



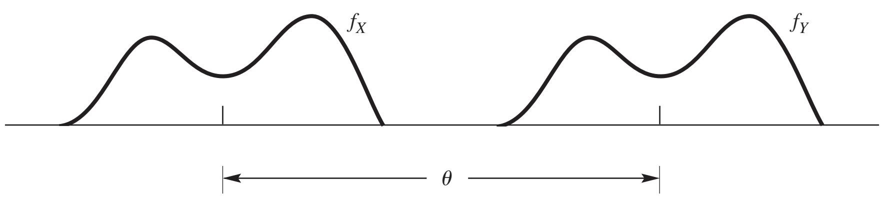
la misma media y la misma varianza; esto es, las distribuciones de población son idénticas. Si H_a es verdadera, entonces $\mu_Y > \mu_X$ y las distribuciones de X_1 y Y_1 son iguales, excepto que el parámetro de localización (μ_Y) para Y_1 es mayor que el parámetro de localización (μ_X) para X_1. En consecuencia, la distribución de Y_1 se desplaza a la derecha de la distribución de X_1 (vea Figura 15.1).

Este es un ejemplo de un *modelo de desplazamiento* (o *localización*) paramétrico de dos muestras. El modelo es paramétrico porque las distribuciones están especificadas (normales) excepto para los valores de los parámetros μ_X, μ_Y y σ^2. La cantidad que la distribución de Y_1 se desplaza a la derecha de la distribución de X_1 es $\mu_Y - \mu_X$ (vea la Figura 15.1). En el resto de esta sección definimos un modelo de desplazamiento que se aplica para cualquier distribución, normal o de otro tipo.

Sea $X_1, X_2, \ldots, X_{n_1}$ una muestra aleatoria de una población con función de distribución $F(x)$ y sea $Y_1, Y_2, \ldots, Y_{n_2}$ una muestra aleatoria de una población con función de distribución $G(y)$. Si deseamos probar que las dos poblaciones tienen la misma distribución, es decir, $H_0: F(z) = G(z)$ contra $H_a: F(z) \neq G(z)$, con la forma real de $F(z)$ y $G(z)$ no especificada, se requiere un método no paramétrico. Observe que H_a es una hipótesis muy general. Muchas veces, un experimentador puede querer considerar la hipótesis alternativa más específica de que Y_1 tiene la misma distribución que X_1 desplazada una cantidad θ (desconocida) (vea Figura 15.2), es decir, que las distribuciones *difieren en localización*. Entonces $G(y) = P(Y_1 \leq y) = P(X_1 \leq y - \theta) = F(y - \theta)$ para algún valor θ de parámetro desconocido. Observe que la forma particular de $F(x)$ continúa no especificada.

En todo este capítulo, si nos referimos al modelo de desplazamiento (localización) de dos muestras, suponemos que $X_1, X_2, \ldots, X_{n_1}$ constituyen una muestra aleatoria de la función de distribución $F(x)$ y que $Y_1, Y_2, \ldots, Y_{n_2}$ constituyen una muestra aleatoria de la función de distribución $G(y) = F(y - \theta)$ para algún valor desconocido θ. Para el modelo de desplazamiento de dos muestras, $H_0: F(z) = G(z)$ es equivalente a $H_0: \theta = 0$. Si θ es mayor (menor) que 0, entonces la distribución de los valores Y está ubicada a la derecha (izquierda) de la distribución de los valores X.

FIGURA **15.2** Dos funciones de densidad, con la densidad para Y desplazada θ unidades a la derecha de la de X



15.3 Prueba de signos para un experimento de observaciones pareadas

Suponga que tenemos n pares de observaciones de la forma (X_i, Y_i) y que deseamos probar la hipótesis de que la distribución de las X es igual a la de las Y, contra la alternativa de que las distribuciones difieren en ubicación (vea Sección 15.2). De manera similar a la que vimos en la Sección 12.3, hagamos $D_i = X_i - Y_i$. Una de las pruebas no paramétricas más sencillas está basada en los signos de estas diferencias y, con justificada razón, se denomina *prueba de signos*. Dada la hipótesis nula de que X_i y Y_i provienen de las mismas distribuciones continuas de probabilidad, la probabilidad de que D_i sea positiva es igual a $1/2$ (como lo es la probabilidad de que D_i sea negativa). Denotemos con M el número total de diferencias positivas (o negativas). Entonces, si las variables X_i y Y_i tienen la misma distribución, M tiene una distribución binomial con $p = 1/2$ y la región de rechazo para una prueba basada en M se puede obtener con el uso de la distribución binomial de probabilidad introducida en el Capítulo 3. La prueba de signos se resume de la siguiente manera.

Prueba de signos para un experimento de observaciones pareadas

Sea $p = P(X > Y)$.

Hipótesis nula: $H_0 : p = 1/2$.

Hipótesis alternativa: $H_a : p > 1/2$ o ($p < 1/2$ o $p \neq 1/2$).

Estadístico de prueba: M = número de diferencias positivas donde $D_i = X_i - Y_i$.

Región de rechazo: para $H_a : p > 1/2$, rechazar H_0 para los valores más grandes de M; para $H_a : p < 1/2$, rechazar H_0 para los valores más pequeños de M; para $H_a : p \neq 1/2$, rechazar H_0 para valores muy grandes o muy pequeños de M.

Suposiciones: los pares (X_i, Y_i) se seleccionan en forma aleatoria e independiente.

El siguiente ejemplo ilustra el uso de la prueba de signos.

EJEMPLO **15.1** El número de fusibles eléctricos defectuosos producidos por cada una de dos líneas de producción, A y B, se registró diariamente durante un periodo de 10 días y se obtuvieron los resultados de la Tabla 15.1. Suponga que ambas líneas produjeron la misma cantidad todos los días. Compare el número de piezas defectuosas producidas por A y B cada día y sea M igual al número de días cuando A rebasó a B. ¿Los datos presentan suficiente evidencia para indicar que cualquiera de las líneas produce más piezas defectuosas que la otra? Exprese la hipótesis nula a probar y use M como un estadístico de prueba.

Solución Forme pares de las observaciones como aparecen en la tabulación de datos y considere M como el número de días en que el número observado de piezas defectuosas de la línea A es mayor que las de la línea B. Dada la hipótesis nula de que las dos distribuciones de piezas defectuosas son idénticas, la probabilidad p de que A exceda a B para un par dado es $p = .5$, dado que no hay empates. En consecuencia, la hipótesis nula es equivalente a la hipótesis de que el parámetro binomial es $p = .5$.

Tabla 15.1 Datos para el ejemplo 15.1

Día	A	B
1	172	201
2	165	179
3	206	159
4	184	192
5	174	177
6	142	170
7	190	182
8	169	179
9	161	169
10	200	210

Valores muy grandes o muy pequeños de M contradicen la hipótesis nula. Por tanto, la región de rechazo para la prueba se ubicará al incluir los valores extremos de M que al mismo tiempo dan un valor de α que es apropiado para la prueba.

Suponga que nos gustaría que el valor de α fuera del orden de .05 o .10. Comenzamos la selección de la región de rechazo al incluir $M = 0$ y $M = 10$ y calcular la α asociada con esta región usando $p(y)$, la distribución de probabilidad para la variable binomial aleatoria (vea el Capítulo 3). Con $n = 10$, $p = .5$, tenemos

$$\alpha = p(0) + p(10) = \binom{10}{0}(.5)^{10} + \binom{10}{10}(.5)^{10} = .002.$$

Como este valor de α es demasiado pequeño, la región se expandirá si se incluye el siguiente par de valores M que se oponen a la hipótesis nula, $M = 1$ y $M = 9$. El valor de α para esta región ($M = 0, 1, 9, 10$) se puede obtener de la Tabla 1, Apéndice 3:

$$\alpha = p(0) + p(1) + p(9) + p(10) = .022.$$

Éste también es demasiado pequeño, de modo que expandimos la región otra vez para incluir $M = 0, 1, 2, 8, 9, 10$. Se puede verificar que el correspondiente valor de α es .11. Suponga que este valor de α es aceptable para el experimentador; entonces empleamos $M = 0, 1, 2, 8, 9, 10$ como la región de rechazo para la prueba.

De los datos, observamos que $m = 2$, de modo que rechazamos la hipótesis nula. Concluimos que existe suficiente evidencia para indicar que las distribuciones poblacionales de las cantidades de fusibles defectuosos no son idénticas. La probabilidad de rechazar la hipótesis nula cuando es verdadera es sólo $\alpha = .11$ y por tanto tenemos una confianza razonable en nuestra conclusión.

El experimentador en este ejemplo está usando el procedimiento de prueba como herramienta burda para detectar líneas de producción con falla. No es probable que el valor hasta cierto punto grande de α lo altere porque él puede fácilmente recolectar datos adicionales si está preocupado por cometer un error tipo I en su conclusión. ■

Los niveles de significancia alcanzados (valores p) para la prueba de signos se calculan como se indica en la Sección 10.6. Específicamente, si $n = 15$ y deseamos probar $H_0 : p =$

$1/2$ contra $H_a : p < 1/2$ con base en el valor observado de $M = 3$, la Tabla 1 del Apéndice 3 se puede usar para determinar que (como $n = 15$, $p = 1/2$)

$$\text{valor } p = P(M \leq 3) = .018.$$

Para la prueba de dos colas ($H_a : p \neq 1/2$), valor $p = 2(.018) = .036$.

EJEMPLO **15.2** Encuentre el valor p asociado con la prueba de signo efectuada en el Ejemplo 15.1.

Solución La prueba del Ejemplo 15.1 es una prueba de dos colas de $H_0 : p = 1/2$ contra $H_a : p \neq 1/2$. El valor calculado de M es $m = 2$, de modo que el valor p es $2P(M \leq 2)$. De acuerdo con la hipótesis nula, M tiene una distribución binomial con $n = 10$, $p = .5$ y la Tabla 1, Apéndice 3, da como resultado

$$\text{valor } p = 2P(M \leq 2) = 2(.055) = .11.$$

Entonces, .11 es el valor *más pequeño* de α para el cual la hipótesis nula se puede rechazar. Observe que el método del valor p da como resultado la misma decisión a la que se llegó en el Ejemplo 15.1 donde se utilizó una prueba formal de nivel $\alpha = .11$. No obstante, el método del valor p eliminó la necesidad de intentar varias regiones de rechazo hasta encontrar una con un valor satisfactorio para α. ■

Un problema que puede surgir en relación con una prueba de signos es que las observaciones asociadas con uno o más pares pueden ser iguales y por tanto puede resultar en empates. Cuando se presente esta situación, borre los pares empatados y reduzca n, el número total de pares.

Usted también hallará situaciones donde n, el número de pares, es grande. Entonces, los valores de α asociados con la prueba de signos pueden calcularse si se usa la aproximación normal a la distribución binomial de probabilidad que estudiamos en la Sección 7.5. Es posible verificar (al comparar probabilidades exactas con sus aproximaciones) que estas aproximaciones serán bastante adecuadas para n de valor muy pequeño, 10 o 15. Este resultado se debe a la simetría de la distribución binomial de probabilidad para $p = .5$. Para $n \geq 25$, la prueba Z del Capítulo 10 será suficiente, donde

$$Z = \frac{M - np}{\sqrt{npq}} = \frac{M - n/2}{(1/2)\sqrt{n}}.$$

Este estadístico se usaría para probar la hipótesis nula $p = .5$ contra la alternativa $p \neq .5$ para una prueba de dos colas o contra la alternativa $p > .5$ (o $p < .5$) para una prueba de una cola. Las pruebas usarían las conocidas regiones de rechazo del Capítulo 10.

Los datos del Ejemplo 15.1 son el resultado de un experimento de observaciones pareadas. Suponga que las diferencias de éstas se distribuyen normalmente con una varianza común σ^2. ¿La prueba de signos detectará un desplazamiento en la localización de las dos poblaciones, en forma tan eficiente como la prueba t de Student? Intuitivamente sospecharíamos que la respuesta es negativa y esto es correcto porque la prueba t de Student usa comparativamente más información. Además de dar el signo de la diferencia, la prueba t usa las magnitudes de las observaciones para obtener valores más precisos para medias muestrales y varianzas. Entonces, podemos decir que la prueba de signos no es tan “eficiente” como la prueba t de Student; pero este enunciado es significativo sólo si las poblaciones se ajustan a la suposición que acabamos de expresar: las diferencias en observaciones pareadas están distribuidas normalmente con una varianza común σ_D^2. La prueba de signos podría ser más eficiente cuando estas suposiciones no se satisfacen.

Prueba de signos para muestras grandes: $n > 25$

Hipótesis nula: $H_0 : p = .5$ (ninguno de estos tratamientos se prefiere al otro).

Hipótesis alternativa: $H_a : p \neq .5$ para una prueba de dos colas (*Nota*: usamos la prueba de dos colas para un ejemplo. Muchos análisis requieren una prueba de una cola).

Estadístico de prueba: $Z = [M - n/2]/[(1/2)\sqrt{n}]$.

Región de rechazo: rechazar H_0 si $z \geq z_{\alpha/2}$ o si $z \leq -z_{\alpha/2}$, donde $z_{\alpha/2}$ se obtiene de la Tabla 3, Apéndice 3.

La prueba de signos en realidad prueba la hipótesis nula de que la *mediana* de las variables D_i es cero contra la alternativa de que es diferente de cero. [Si la mediana de las variables D_i es cero, implica que $P(D_i < 0) = P(D_i > 0)$.] Si las variables X_i y Y_i tienen la misma distribución, la mediana de las variables D_i será cero, como ya dijimos antes. No obstante, para modelos que no sean el modelo de desplazamiento hay otras situaciones en las que la mediana de las variables D_i es cero. En estos casos la hipótesis nula para la prueba de signos es ligeramente más general que el enunciado de que X_i y Y_i tienen la misma distribución.

Resumiendo, la prueba de signos es un procedimiento no paramétrico que se aplica con facilidad para comparar dos poblaciones. No se hacen suposiciones respecto a las distribuciones poblacionales que sirven de base. El valor del estadístico de prueba se puede obtener rápidamente por conteo visual y la región de rechazo (o valor p) se puede hallar sin problema si se usa una tabla de probabilidades binomiales. Además, no es necesario conocer los valores exactos de pares de respuestas, sólo si $X_i > Y_i$ para cada par (X_i, Y_i). El Ejercicio 15.5 muestra un ejemplo del uso de la prueba de signos para datos de este tipo.

Ejercicios

15.1 ¿Qué niveles de significancia entre $\alpha = .01$ y $\alpha = .15$ hay para una prueba de signos de dos colas con 25 observaciones pareadas? (Haga uso de valores tabulados en la Tabla 1, Apéndice 3, $n = 25$.) ¿Cuáles son las correspondientes regiones de rechazo?

15.2 Un estudio publicado en la *American Journal of Public Health* (*Science News*) —el primero en seguir niveles de plomo en la sangre de aficionados a portar armas pero que son respetuosos de las leyes y usan polígonos de tiro bajo techo— demuestra un considerable riesgo de envenenamiento por plomo.[1] Las mediciones de exposición al plomo se hicieron en 17 miembros de un grupo de estudiantes a oficiales de policía antes, durante y después de un periodo de 3 meses de instrucción sobre manejo de armas de fuego en un polígono de tiro bajo techo propiedad del gobierno. Ninguno de los alumnos tenía niveles elevados de plomo en la sangre antes del curso, pero 15 de los 17 terminaron el curso con niveles de plomo en la sangre considerados como “elevados” por la Occupational Safety and Health Administration (OSHA). ¿Hay suficiente evidencia para decir que el polígono de tiro en interiores aumenta las lecturas del nivel de sangre?

1. *Fuente: Science News*, 136 (agosto de 1989): 126.

a Proporcione el valor p asociado.

b ¿Qué se concluiría con un nivel de significancia $\alpha = .01$?

c Use la aproximación normal para dar el valor p aproximado. ¿La aproximación parece ser adecuada cuando $n = 17$?

15.3 Datos clínicos relacionados con la efectividad de dos medicamentos para tratar una enfermedad se recolectaron de diez hospitales. El número de pacientes tratados con los medicamentos difirió para los diversos hospitales. Los datos se dan en la siguiente tabla.

	Medicamento A			Medicamento B		
Hospital	Número tratado	Número de recuperados	Porcentaje de recuperados	Número tratado	Número de recuperados	Porcentaje de recuperados
1	84	63	75.0	96	82	85.4
2	63	44	69.8	83	69	83.1
3	56	48	85.7	91	73	80.2
4	77	57	74.0	47	35	74.5
5	29	20	69.0	60	42	70.0
6	48	40	83.3	27	22	81.5
7	61	42	68.9	69	52	75.4
8	45	35	77.8	72	57	79.2
9	79	57	72.2	89	76	85.4
10	62	48	77.4	46	37	80.4

a ¿Los datos indican una diferencia en los *porcentajes de recuperación* para los dos medicamentos? Proporcione el valor p asociado.

b ¿Por qué podría ser inapropiado usar la prueba t para analizar los datos?

15.4 Para una comparación de la efectividad académica de dos escuelas secundarias A y B se diseñó un experimento usando diez parejas de mellizos idénticos, cada uno de los cuales acababa de terminar el sexto año escolar. En cada caso, los mellizos de la misma pareja habían obtenido su escolaridad previa en los mismos salones de clase en cada nivel escolar. Un niño fue seleccionado aleatoriamente de cada pareja y asignado a la escuela A; el otro fue enviado a la escuela B. Cerca ya de terminar el noveno grado, se aplicó un examen de conocimientos a cada niño del experimento. Los resultados se muestran en la siguiente tabla.

Par de mellizos	A	B	Par de mellizos	A	B
1	67	39	6	50	52
2	80	75	7	63	56
3	65	69	8	81	72
4	70	55	9	86	89
5	86	74	10	60	47

a Usando la prueba de signos, pruebe la hipótesis de que las dos escuelas son iguales en efectividad académica, según se mide con las calificaciones del examen de conocimientos, contra la alternativa de que las escuelas no son igualmente eficientes. Dé el nivel de significancia alcanzado. ¿Qué concluiría usted con $\alpha = .05$?

b Supongamos que se considera que la escuela secundaria A tiene un mejor profesorado y mejores instalaciones para la enseñanza. Pruebe la hipótesis de igual efectividad académica contra la alternativa de que la escuela A es superior. ¿Cuál es el valor p asociado con esta prueba?

15.5 Los productos alimenticios nuevos son sometidos frecuentemente a pruebas de sabor por un panel de jueces, a quienes por lo general se les pide expresen su preferencia por un alimento sobre otro para que no haya necesidad de usar una escala cuantitativa. Suponga que dos nuevas mezclas, A y B, de un refresco de sabor naranja se presentan a diez jueces. Las preferencias de los jueces se proporcionan en la siguiente tabla. ¿Esta evidencia indica una diferencia importante entre los gustos de A y B, con un nivel de significancia de 5%?

Juez	Preferencia	Juez	Preferencia
1	A	6	A
2	A	7	B
3	A	8	A
4	A	9	B
5	A	10	A

15.6 En noches claras y frías de la región productora de cítricos del centro de Florida, la ubicación precisa de las temperaturas abajo del punto de congelación es importante porque los métodos para proteger los árboles contra condiciones de congelamiento son muy costosos. Un método para ubicar probables puntos fríos consiste en relacionar la temperatura con la elevación. Se piensa que en noches en calma los puntos fríos estarán en elevaciones bajas. Los puntos más altos y más bajos de una plantación particular dieron las temperaturas mínimas que se citan en la siguiente tabla, para diez noches frías en un invierno reciente.

Noche	Elevación alta	Elevación baja
1	32.9	31.8
2	33.2	31.9
3	32.0	29.2
4	33.1	33.2
5	33.5	33.0
6	34.6	33.9
7	32.1	31.0
8	33.1	32.5
9	30.2	28.9
10	29.1	28.0

a ¿Hay evidencia suficiente para apoyar la conjetura de que elevaciones bajas tienden a ser más frías? (Use la prueba de signos. Dé el valor p asociado.)

b ¿Sería razonable usar una prueba t en estos datos? ¿Por qué sí o por qué no?

15.7 Se realizó un experimento psicológico para comparar los tiempos de respuesta (en segundos) para dos estímulos diferentes. Para eliminar la variabilidad natural de una persona a otra en las respuestas, ambos estímulos se aplicaron a cada uno de nueve sujetos, permitiendo así un análisis de la diferencia entre los tiempos de respuesta de cada persona. Los resultados se presentan en la siguiente tabla.

Sujeto	Estímulo 1	Estímulo 2
1	9.4	10.3
2	7.8	8.9
3	5.6	4.1
4	12.1	14.7
5	6.9	8.7
6	4.2	7.1
7	8.8	11.3
8	7.7	5.2
9	6.4	7.8

a Use la prueba de signos para determinar si existe suficiente evidencia para indicar una diferencia en la respuesta media para los dos estímulos. Use una región de rechazo para la cual $\alpha \leq .05$.

b Pruebe la hipótesis de que no haya diferencia en la respuesta media, usando una prueba t de Student.

15.8 Consulte el Ejercicio 12.15. Usando la prueba de signos, ¿encuentra suficiente evidencia para apoyar la conclusión de que los tiempos de terminación difieren para las dos poblaciones? Use $\alpha = .10$.

15.9 El conjunto de datos de la siguiente tabla representa el número de accidentes industriales en 12 plantas de manufactura, durante periodos de una semana antes y después de intensa promoción sobre seguridad.

Planta	Antes	Después	Planta	Antes	Después
1	3	2	7	5	3
2	4	1	8	3	3
3	6	3	9	2	0
4	3	5	10	4	3
5	4	4	11	4	1
6	5	2	12	5	2

a ¿Los datos apoyan la afirmación de que la campaña fue un éxito? ¿Cuál es el nivel de significancia alcanzado? ¿Qué concluiría usted con $\alpha = .01$?

b Analice los problemas asociados con un análisis paramétrico diseñado para contestar la pregunta del inciso a.

15.4 Prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas

Al igual que en la Sección 15.3 suponga que tenemos n observaciones pareadas de la forma (X_i, Y_i) y que $D_i = X_i - Y_i$. De nuevo suponemos que estamos interesados en probar la hipótesis de que las X y las Y tienen la misma distribución contra la alternativa de que las distribuciones difieren en localización. De acuerdo con la hipótesis nula de que no hay diferencia en las distribuciones de las X y las Y, se esperaría (en promedio) que la mitad de las diferencias en los pares sean negativas y la mitad positivas. Esto es, el número esperado de diferencias negativas entre pares es $n/2$ (donde n es el número de pares). Además, se podría inferir que las diferencias positivas y negativas de igual magnitud absoluta deberían presentarse con la misma probabilidad. Si ordenáramos las diferencias de acuerdo con sus valores absolutos y las clasificáramos de menor a mayor, las sumas de los rangos esperados para las diferencias negativas y positivas serían iguales. Las diferencias grandes en las sumas de los rangos asignados a las diferencias positivas y negativas darían evidencia para indicar un desplazamiento de localización entre las dos distribuciones.

Para llevar a cabo la prueba de Wilcoxon calculamos las diferencias (D_i) para cada uno de los n pares. Las diferencias iguales a cero se eliminan y el número de pares, n, se reduce de conformidad. Entonces clasificamos los *valores absolutos* de las diferencias asignando un 1 al más pequeño, un 2 al segundo más pequeño y así sucesivamente. Si dos o más diferencias absolutas están empatadas para el mismo rango, entonces el promedio de las clasificaciones que se hubieran asignado a estas diferencias se asigna a cada miembro del grupo empatado. Por ejemplo, si dos diferencias absolutas están empatadas para los rangos 3 y 4, entonces cada una recibe clasificaciones 3.5 y a la siguiente diferencia absoluta más alta se le asigna el rango 5. Entonces calculamos la suma de los rangos (suma de rango) para las diferencias negativas y también calculamos la suma de rango para las diferencias positivas. Para una prueba de dos colas usamos T, la *más* *pequeña* de estas dos cantidades, como estadístico de prueba para probar la hipótesis nula de que los dos histogramas de frecuencia relativa poblacionales son idénticos. Cuanto menor sea el valor de T, mayor será el valor de evidencia a favor del rechazo de la hipótesis nula. En consecuencia, rechazaremos la hipótesis nula si T es menor o igual a algún valor, por ejemplo, T_0.

Para detectar la alternativa unilateral, que afirma que la distribución de las X se desplaza a la derecha de la de las Y, usamos la suma de rango T^- de las diferencias negativas y rechazamos la hipótesis nula para valores pequeños de T^-, por ejemplo, $T^- \le T_0$. Si deseamos detectar un desplazamiento de la distribución de las Y a la derecha de las X, usamos la suma de rango T^+ de las diferencias positivas como estadístico de prueba y rechazamos valores pequeños de T^+, por ejemplo, $T^+ \le T_0$.

La probabilidad de que T sea menor o igual a algún valor T_0 se ha calculado para una combinación de tamaños muestrales y valores de T_0. Estas probabilidades, dadas en la Tabla 9, Apéndice 3, se pueden usar para hallar la región de rechazo para la prueba basada en T.

Por ejemplo, supongamos que usted tiene $n = 7$ pares y desea realizar una prueba de dos colas de la hipótesis nula de que las dos distribuciones de frecuencia relativa poblacionales son idénticas. Entonces, con $\alpha = .05$, rechazaría la hipótesis nula para todos los valores de T menores o iguales a 2. La región de rechazo para la prueba Wilcoxon de suma de rango para un experimento pareado es siempre de esta forma: rechazar la hipótesis nula si $T \le T_0$ donde T_0 es el valor crítico para T. Los límites para el nivel de significancia alcanzado (valor p) se determinan como sigue. Para una prueba de dos colas, si $T = 3$ se observa cuando $n = 7$, la Tabla 9, Apéndice 3, indica que H_0 sería rechazada si $\alpha = .1$, pero no si $\alpha = .05$. Entonces, $.05 <$ valor $p < .1$. Para una alternativa unilateral de que las X están desplazadas a la derecha de las Y con $n = 7$ y $\alpha = .05$, H_0 es rechazada si $T = T^- \le 4$. En este caso, si $T = T^- = 1$, entonces $.01 <$ valor $p < .025$. La prueba basada en T, llamada *prueba de de rangos con signo Wilcoxon*, se resume como sigue.

Prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas

H_0: las distribuciones poblacionales para las X y las Y son idénticas.

H_a: (1) las dos distribuciones poblacionales difieren en localización (dos colas), *o bien* (2) la distribución de frecuencia relativa poblacional para las X se desplaza a la derecha de la de las Y (una cola).

Estadístico de prueba:

1. Para una prueba de dos colas, use $T = \text{mín}(T^+, T^-)$, donde $T^+ =$ suma de los rangos de las diferencias positivas y $T^- =$ suma de los rangos de las diferencias negativas.
2. Para una prueba de una cola (para detectar la alternativa de una cola que acabamos de dar), use la suma de rango T^- de las diferencias negativas.[2]

Región de rechazo:

1. Para una prueba de dos colas, rechace H_0 si $T \le T_0$, donde T_0 es el valor crítico para la prueba de dos lados dada en la Tabla 9, Apéndice 3.
2. Para una prueba de una cola (como se describe líneas antes), rechace H_0 si $T^- \le T_0$, donde T_0 es el valor crítico para la prueba unilateral.

2. Para detectar un desplazamiento de la distribución de las Y a la derecha de la distribución de las X, use la suma de rango T^+, la suma de los rangos de las diferencias positivas, y rechace H_0 si $T^+ \le I_0$.

EJEMPLO **15.3** Debido a la variación de un horno a otro, se utilizó un experimento de observaciones pareadas para probar diferencias en pasteles elaborados usando la mezcla A y la mezcla B. Dos pasteles elaborados usando cada mezcla, se hornearon en cada uno de seis hornos diferentes (un total de 12 pasteles). Pruebe la hipótesis de que no hay diferencia en las distribuciones poblacionales de las densidades de pastel usando las dos mezclas. ¿Qué se puede decir acerca del nivel de significancia alcanzado?

Solución Los datos originales y las diferencias en densidades (en onzas por pulgada cúbica) para los seis pares de pasteles se muestran en la Tabla 15.2.

Al igual que con otras pruebas no paramétricas, la hipótesis nula a probar es que las dos distribuciones de frecuencia poblacionales de densidades de pastel son idénticas. La hipótesis alternativa es que las distribuciones difieren en localización, lo cual implica que se requiere una prueba de dos colas.

Debido a que la cantidad de datos es pequeña, realizaremos nuestra prueba usando $\alpha = .10$. De la Tabla 9, Apéndice 3, el valor crítico de T para una prueba de dos colas, $\alpha = .10$, es $T_0 = 2$. En consecuencia, rechazaremos H_0 si $T \leq 2$.

Hay sólo una diferencia positiva y tiene rango 3; por tanto, $T^+ = 3$. Como $T^+ + T^- = n(n + 1)/2$ (¿por qué?), $T^- = 21 - 3 = 18$ y el valor observado de T es mín(3, 18) = 3. Observe que 3 excede el valor crítico de T, lo cual implica que no hay suficiente evidencia para indicar una diferencia en las dos distribuciones de frecuencia poblacional en las densidades de los pasteles. Como no podemos rechazar H_0 para $\alpha = .10$, sólo podemos decir que valor $p > .10$.

Tabla 15.2 Datos pareados y sus diferencias para el Ejemplo 15.3

A	B	Diferencia A − B	Diferencia absoluta	Rango de diferencia absoluta
.135	.129	.006	.006	3
.102	.120	−.018	.018	5
.108	.112	−.004	.004	1.5
.141	.152	−.011	.011	4
.131	.135	−.004	.004	1.5
.144	.163	−.019	.019	6

■

Aunque la Tabla 9, Apéndice 3, es aplicable para valores de n (el número de pares de datos) de hasta $n = 50$, es conveniente observar que T^+ (o T^-) estará distribuida normalmente en forma aproximada cuando la hipótesis nula sea verdadera y n sea grande (25 o más, por ejemplo). Esto hace posible que construyamos una prueba Z de muestra grande, donde si $T = T^+$,

$$E(T^+) = \frac{n(n+1)}{4} \quad \text{y} \quad V(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

Entonces el estadístico Z

$$Z = \frac{T^+ - E(T^+)}{\sqrt{V(T^+)}} = \frac{T^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$$

se puede usar como estadístico de prueba. Entonces, para una prueba de dos colas y $\alpha = .05$, rechazaríamos la hipótesis de distribuciones poblacionales idénticas cuando $|z| \geq 1.96$. Para una prueba de una cola de que la distribución de las X está desplazada a la derecha (izquierda) de la distribución de las Y, rechace H_0 cuando $z > z_\alpha$ $(z < -z_\alpha)$.

Una prueba de rangos con signo de Wilcoxon con muestras grandes para un experimento de observaciones pareadas: $n > 25$

Hipótesis nula: H_0: las distribuciones de frecuencia relativa poblacionales para las X y las Y son idénticas.

Hipótesis alternativa: (1) H_a: las dos distribuciones de frecuencia relativa poblacionales difieren en localización (una prueba de dos colas),

o bien, (2) la distribución de frecuencia relativa poblacional para las X está desplazada a la derecha (o izquierda) de la distribución de frecuencia relativa de las Y (pruebas de una cola).

Estadístico de prueba $Z = \dfrac{T^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$.

Región de rechazo: rechazar H_0 si $z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ para una prueba de dos colas. Para detectar un desplazamiento en las distribuciones de las X a la derecha de las Y, rechazar H_0 cuando $z \geq z_\alpha$. Para detectar un desplazamiento en la dirección opuesta, rechazar H_0 si $z \leq -z_\alpha$.

Ejercicios

15.10 Se lleva a cabo un experimento de observaciones pareadas que usa n pares de observaciones, si T^+ = la suma de las columnas de los valores absolutos de las diferencias positivas y T^- = la suma de las columnas de los valores absolutos de las diferencias negativas, ¿por qué es $T^+ + T^- = n(n+1)/2$?

15.11 Consulte el Ejercicio 15.10. Si T^+ se ha calculado, ¿cuál es la forma más fácil de determinar el valor de T^-? Si $T^+ > n(n+1)/4$, ¿es $T = T^+$ o T^-? ¿Por qué?

15.12 La siguiente tabla muestra las calificaciones de un grupo de 15 estudiantes en matemáticas y artes.

Estudiante	Matemáticas	Artes	Estudiante	Matemáticas	Artes
1	22	53	9	62	55
2	37	68	10	65	74
3	36	42	11	66	68
4	38	49	12	56	64
5	42	51	13	66	67
6	58	65	14	67	73
7	58	51	15	62	65
8	60	71			

a Use la prueba de rangos con signo de Wilcoxon para determinar si las localizaciones de las distribuciones de calificaciones para estos estudiantes difieren de manera importante para las dos materias. Establezca límites para el valor p e indique la conclusión apropiada cuando $\alpha = .05$.

b Exprese las hipótesis nula y alternativa para la prueba que realizó en el inciso a.

15.13 Consulte el Ejercicio 15.4. ¿Qué respuestas se obtienen si la prueba de rangos con signo de Wilcoxon se usa en el análisis de los datos? Compare estas respuestas con las obtenidas en el Ejercicio 15.4.

15.14 Consulte el Ejercicio 15.6(a). Conteste la pregunta usando la prueba de rangos con signo de Wilcoxon.

15.15 A ocho personas se les pidió ejecutaran un sencillo trabajo de ensamblar un rompecabezas en condiciones normales y en condiciones de estrés. Durante la condición de estrés, a las personas se les indicó que se les aplicaría una ligera descarga eléctrica 3 minutos después de empezar el experimento y cada 30 segundos de ahí en adelante hasta terminar el trabajo. Las lecturas de presión sanguínea se tomaron en ambas condiciones. Los datos de la siguiente tabla representan la lectura más alta durante el experimento.

Sujeto	Normal	Estrés
1	126	130
2	117	118
3	115	125
4	118	120
5	118	121
6	128	125
7	125	130
8	120	120

¿Los datos presentan suficiente evidencia para indicar lecturas de presión sanguínea más alta durante los condiciones de estrés? Analice los datos usando la prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas. Proporcione el valor p apropiado.

15.16 Se emplearon dos métodos, A y B, para controlar el tránsito en cada uno de $n = 12$ cruceros durante una semana. Los números de accidentes que ocurrieron durante este tiempo se registraron en la siguiente tabla. El orden de uso (cuál método se empleó para la primera semana) se eligió aleatoriamente para cada crucero.

Crucero	Método A	Método B	Crucero	Método A	Método B
1	5	4	7	2	3
2	6	4	8	4	1
3	8	9	9	7	9
4	3	2	10	5	2
5	6	3	11	6	5
6	1	0	12	1	1

a Analice estos datos usando la prueba de signos.

b Analice estos datos usando la prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas.

15.17 Investigadores en odontología han desarrollado un nuevo material para prevenir la caries, un sellador plástico que se aplica a las superficies de masticación de los dientes. Para determinar si el sellador es eficaz, se aplicó a la mitad de los dientes de cada uno de 12 niños en edad escolar. Después de 2 años se hizo un recuento del número de caries en los dientes con sellador y en los dientes no tratados. Los resultados se muestran en la siguiente tabla. ¿Hay suficiente evidencia para indicar que los dientes con sellador son menos propensos a la caries que los dientes no tratados? Pruebe usando $\alpha = 0.05$.

Niño	Con sellador	No tratado	Niño	Con sellador	No tratado
1	3	3	7	1	5
2	1	3	8	2	0
3	0	2	9	1	6
4	4	5	10	0	0
5	1	0	11	0	3
6	0	1	12	4	3

15.18 Consulte el Ejercicio 12.16. Con $\alpha = .01$, use la prueba de rangos con signo de Wilcoxon para ver si hubo una pérdida importante en la profundidad del humus entre el principio y el fin del estudio.

15.19 Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una función de distribución continua $F(y)$. Se desea probar una hipótesis respecto a la mediana ξ de $F(y)$. Construya una prueba de $H_0 : \xi = \xi_0$ contra $H_a : \xi \neq \xi_0$, donde ξ_0 es una constante especificada.

a Use la prueba de signos.

b Use la prueba de rangos con signo de Wilcoxon.

15.20 El vocero de una organización que apoya reducciones en impuesto a la propiedad, de cierta sección de una ciudad, expresó que el ingreso medio anual de los jefes de familia en esa sección era de \$15 000. Una muestra aleatoria de diez jefes de familia de esa sección reveló los siguientes ingresos anuales:

14 800	16 900	18 000	19 100	13 200
18 500	20 000	19 200	15 100	16 500

Con $\alpha = .10$, pruebe la hipótesis de que el ingreso medio para la población de esa sección es \$15 000 contra la alternativa de que sea mayor que \$15 000.

a Use la prueba de signos.

b Use la prueba de rangos con signo de Wilcoxon.

15.5 Uso de rangos para comparar dos distribuciones poblacionales: muestras aleatorias independientes

Una prueba estadística para comparar dos poblaciones basada en muestras aleatorias independientes, la prueba de *suma de rangos*, fue propuesta por Frank Wilcoxon en 1945. De nuevo suponemos que estamos interesados en probar si las dos poblaciones tienen la misma distribución contra el desplazamiento (o localización) alternativo (véase la Sección 15.2). Supongamos que usted debe seleccionar muestras aleatorias independientes de n_1 y n_2 observaciones de las poblaciones I y II, respectivamente. La idea de Wilcoxon era combinar las $n_1 + n_2 = n$ observaciones y clasificarlas, en orden de magnitud, de 1 (la más pequeña) a n (la más grande). Los empates se tratan como en la Sección 15.4. Esto es, si dos o más observaciones están empatadas para el mismo rango, el promedio de los rangos que se hubiera asignado a estas observaciones se asigna a cada miembro del grupo empatado. Si las observaciones se seleccionaron de poblaciones idénticas, las *sumas de rango* para las muestras deben ser más o menos proporcionales a los tamaños muestrales n_1 y n_2. Por ejemplo, si n_1 y n_2 fueran iguales, se esperaría que las sumas de rango fueran casi iguales. En contraste, si las observaciones en una población, la I por ejemplo, tendían a ser mayores que las de la población II, las observaciones de la muestra I tenderían a recibir las clasificaciones más altas y la muestra I tendría una suma de rango mayor que lo esperado. Entonces (con los tamaños muestrales siendo iguales), si una suma de rango es muy grande (y, de modo correspondiente, la otra es muy pequeña), esto puede indicar una diferencia importante, desde el punto de vista estadístico, entre las localizaciones de las dos poblaciones.

Mann y Whitney propusieron en 1947 una prueba estadística equivalente que también empleaba sumas de rango para dos muestras. Como la prueba U de Mann–Whitney y las tablas de valores críticos de U se presentan con tanta frecuencia en la literatura, explicaremos su uso en la Sección 15.6 y daremos varios ejemplos de sus aplicaciones. En esta sección ilustramos la lógica de la prueba de la suma de rango y demostramos cómo determinar la región de rechazo para la prueba y el valor de α.

EJEMPLO **15.4** La cantidad de bacterias por unidad de volumen se muestran en la Tabla 15.3 para dos tipos de cultivos, I y II; se hicieron cuatro observaciones para cada cultivo. Con n_1 y n_2 represente el número de observaciones en las muestras I y II, respectivamente.

Para los datos dados en la Tabla 15.3, los rangos correspondientes se muestran en la Tabla 15.4. ¿Estos datos presentan suficiente evidencia para indicar una diferencia en la localización de las distribuciones poblacionales para los cultivos I y II?

Tabla 15.3 Datos para el Ejemplo 15.4

I	II
27	32
31	29
26	35
25	28

Solución Sea W igual a la suma de rango para la muestra I (para esta muestra, $W = 12$). Ciertamente, valores muy pequeños o muy grandes de W dan evidencia para indicar una diferencia entre las localizaciones de las dos distribuciones poblacionales; de aquí que W, la *suma de rango*, se pueda emplear como estadístico de prueba.

La región de rechazo para una prueba determinada se obtiene en la misma forma que para la prueba de signos. Empezamos por seleccionar los valores de W más contradictorios como la región de rechazo y sumamos éstos hasta que α sea de tamaño aceptable.

Tabla 15.4 Rangos

	I	II
	3	7
	6	5
	2	8
	1	4
Suma de rango	12	24

La suma de rangos mínima incluye los rangos 1, 2, 3, 4 o $W = 10$. Del mismo modo, la máxima incluye los rangos 5, 6, 7, 8, con $W = 26$. Por tanto, incluimos estos dos valores de W en la región de rechazo. ¿Cuál es el valor correspondiente de α?

Hallar el valor de α es un problema de probabilidad que se puede resolver usando los métodos del Capítulo 2. Si las poblaciones son idénticas, toda permutación de los 8 rangos representa un punto muestral y es igualmente probable. Entonces, α es la suma de las probabilidades de los mismos puntos (arreglos) que implican $W = 10$ o $W = 26$. El número total de permutaciones de los ocho rangos es 8! El número de arreglos diferentes de los rangos 1, 2, 3, 4 en la muestra I con los 5, 6, 7, 8 de la muestra II es $4! \times 4!$. De igual manera, el número de arreglos que ponen el valor máximo de W en la muestra I (rangos 5, 6, 7, 8) es $4! \times 4!$. Entonces, la probabilidad de que $W = 10$ o $W = 26$ es

$$p(10) + p(26) = \frac{(2)(4!)(4!)}{8!} = \frac{2}{\binom{8}{4}} = \frac{1}{35} = .029\,.$$

Si este valor de α es demasiado pequeño, la región de rechazo se puede agrandar para incluir las siguientes sumas de rangos más pequeña y más grande, $W = 11$ y $W = 25$. La suma de rango $W = 11$ incluye los rangos 1, 2, 3, 5, y

$$p(11) = \frac{4!\,4!}{8!} = \frac{1}{70}.$$

Del mismo modo,

$$p(25) = \frac{1}{70}\,.$$

Entonces,

$$\alpha = p(10) + p(11) + p(25) + p(26) = \frac{2}{35} = .057\,.$$

La expansión de la región de rechazo para incluir 12 y 24 aumenta considerablemente el valor de α. El conjunto de puntos muestrales que da un rango de 12 incluye todos los puntos muestrales asociados con clasificaciones de (1, 2, 3, 6) y (1, 2, 4, 5). Así,

$$p(12) = \frac{(2)(4!)(4!)}{8!} = \frac{1}{35},$$

y

$$\begin{aligned}\alpha &= p(10) + p(11) + p(12) + p(24) + p(25) + p(26)\\ &= \frac{1}{70} + \frac{1}{70} + \frac{1}{35} + \frac{1}{35} + \frac{1}{70} + \frac{1}{70} = \frac{4}{35} = .114.\end{aligned}$$

Este valor de α podría ser considerado demasiado grande para fines prácticos. Por tanto, estamos más satisfechos con la región de rechazo $W = 10, 11, 25$ y 26.

La suma de rangos para la muestra, $W = 12$, no cae en esta región de rechazo preferida, de modo que no tenemos evidencia suficiente para rechazar la hipótesis de que las distribuciones poblacionales de la cantidad de bacterias para los dos cultivos sean idénticas. ■

15.6 Prueba U de Mann–Whitney: muestras aleatorias independientes

El estadístico U de Mann–Whitney se obtiene al ordenar todas las $(n_1 + n_2)$ observaciones según su magnitud y al contar el número de observaciones en la muestra I que precede a cada observación de la muestra II. El estadístico U es la suma de estas cantidades. En el resto de esta sección denotamos las observaciones en la muestra I como $x_1, x_2, \ldots, x_{n_1}$ y las observaciones en la muestra II como $y_1, y_2, \ldots, y_{n_2}$.

Por ejemplo, las ocho observaciones ordenadas del Ejemplo 15.4 son

25	26	27	28	29	31	32	35
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$y_{(1)}$	$y_{(2)}$	$x_{(4)}$	$y_{(3)}$	$y_{(4)}$

La observación y más pequeña es $y_{(1)} = 28$, y la preceden $u_1 = 3$ de la muestra x. Del mismo modo, $u_2 = 3$ observaciones de las x preceden a $y_{(2)} = 29$ y $u_3 = 4$, y $u_4 = 4$ observaciones de las x preceden a $y_{(3)} = 32$ y $y_{(4)} = 35$, respectivamente. Entonces,

$$U = u_1 + u_2 + u_3 + u_4 = 3 + 3 + 4 + 4 = 14.$$

Valores muy grandes o muy pequeños de U implican una separación de las x y las y ordenadas, en consecuencia demuestran que existe una diferencia (un desplazamiento de localización) entre las distribuciones de las poblaciones I y II.

Como se indica en la Sección 15.5, el estadístico U de Mann–Whitney está relacionado con la suma de rangos de Wilcoxon. De hecho, se puede demostrar (Ejercicio 15.75) que

Fórmula para el estadístico U de Mann–Whitney

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W,$$

donde n_1 = número de observaciones en la muestra I,
n_2 = número de observaciones en la muestra II,
W = suma de rango para la muestra I.

Como se puede ver de la fórmula para U, U es pequeña cuando W es grande, lo que es probable que ocurra cuando la distribución de la población I se desplaza a la derecha de la distribución de la población II. En consecuencia, para realizar una prueba de una cola para detectar un desplazamiento en la distribución de la población I a la derecha de la distribución de la población II, usted rechazará la hipótesis nula de no diferencia en las distribuciones poblacionales si $U \leq U_0$, donde $\alpha = P(U \leq U_0)$ es de tamaño apropiado.

Algunos resultados útiles acerca de la distribución de U:

1. Los posibles valores de U son $0, 1, 2, \ldots, n_1 n_2$.
2. La distribución de U es simétrica alrededor de $(n_1 n_2)/2$. Esto es, para cualquier $a > 0$, $P[U \leq (n_1 n_2)/2 - a] = P[U \geq (n_1 n_2)/2 + a]$.
3. El resultado en (2) implica que $P(U \leq U_0) = P(U \geq n_1 n_2 - U_0)$.

Si se desea realizar una prueba de una cola para detectar un desplazamiento en la distribución de la población I a la izquierda de la distribución de la población II, se rechazaría H_0 si U es muy grande, específicamente si $U \geq n_1n_2 - U_0$, donde U_0 es tal que $\alpha = P(U \geq n_1n_2 - U_0) = P(U \leq U_0)$ es de tamaño aceptable.

La Tabla 8, Apéndice 3, proporciona la probabilidad de que un valor observado de U sea menor que varios valores, U_0. Éste es el valor de α para una prueba de una cola. Para realizar una prueba de dos colas, es decir, para detectar diferencia en las localizaciones de las poblaciones I y II, rechazar H_0 si $U \leq U_0$ o $U \geq n_1n_2 - U_0$, donde $P(U \leq U_0) = \alpha/2$.

Para saber cómo localizar la región de rechazo para la prueba U de Mann–Whitney, suponga que $n_1 = 4$ y $n_2 = 5$. Entonces, se consultaría la tercera sección de la Tabla 8, Apéndice 3 (la correspondiente a $n_2 = 5$). Observe que la tabla está construida suponiendo que $n_1 \leq n_2$. Esto es, usted debe identificar siempre la muestra más pequeña como muestra I. De la tabla vemos, por ejemplo, $P(U \leq 2) = .0317$ y $P(U \leq 3) = .0556$. Por tanto, si se desea realizar una prueba U de Mann–Whitney de cola inferior con $n_1 = 4$ y $n_2 = 5$ para α cercana a .05, se debe rechazar la hipótesis nula de igualdad de distribuciones de frecuencia relativa poblacional cuando $U \leq 3$. La probabilidad de un error tipo I para la prueba es $\alpha = .0556$.

Cuando se aplique la prueba a un conjunto de datos, es posible hallar que algunas de las observaciones son de igual valor. Los empates en las observaciones se pueden manejar al promediar los rangos que se hubieran asignado a las observaciones empatadas y asignar este promedio a cada una. De este modo, si tres observaciones están empatadas y han de recibir rangos 3, 4 y 5, asignamos el rango 4 a todas ellas. La siguiente observación de la secuencia recibe rango 6 y los rangos 3 y 5 no aparecen. De manera similar, si dos observaciones están empatadas para los rangos 3 y 4, cada una recibe rangos 3.5 y los rango 3 y 4 no aparecen.

La Tabla 8, Apéndice 3, se puede usar también para hallar el nivel de significancia observado para una prueba. Por ejemplo, si $n_1 = 5$, $n_2 = 5$ y $U = 4$, el valor p para una prueba de una cola de que la distribución de la población I se desplaza a la derecha de la distribución de la población II es

$$P\{U \leq 4\} = .0476.$$

Si la prueba es de dos colas, el valor p es

$$2(.0476), \text{ o sea } .0952.$$

La prueba *U* de Mann–Whitney

La población I es la población de la cual se tomó la muestra más pequeña.

Hipótesis nula: H_0 : las distribuciones de las poblaciones I y II son idénticas.
Hipótesis alternativa: (1) H_a : las distribuciones de las poblaciones I y II tienen localizaciones diferentes (una prueba de dos colas),
o bien, (2) la distribución de la población I se desplaza a la derecha de la distribución de la población II, *o* (3) la distribución de la población I se desplaza a la izquierda de la distribución de la población II.
Estadístico de prueba: $U = n_1n_2 + [n_1(n_1 + 1)]/2 - W$.
Región de rechazo: (1) para la prueba de dos colas y un valor dado de α, rechazar H_0 si $U \leq U_0$ o $U \geq n_1n_2 - U_0$, donde $P(U \leq U_0) = \alpha/2$. [*Nota*: observe que U_0 es el valor tal que $P(U \leq U_0)$ es igual a la mitad de α.]
(2) Para probar que la población I está desplazada a la derecha de la población II con un valor determinado de α, rechazar H_0 si $U \leq U_0$, donde $P(U \leq U_0) = \alpha$.

(3) Para probar que la población I está desplazada a la izquierda de la población II con un valor determinado de α, rechazar H_0 si $U \geq n_1n_2 - U_0$, donde $P(U \leq U_0) = \alpha$.

Suposiciones: las muestras se han seleccionado aleatoriamente y de manera independiente de sus poblaciones respectivas. Los empates en las observaciones se pueden manejar al promediar los rangos que hubieran sido asignados a las observaciones empatadas y asignar este rango promedio a cada una. Así, si tres observaciones están empatadas y han de recibir rangos 3, 4 y 5, asignamos un rango 4 a todas ellas.

EJEMPLO **15.5** Pruebe la hipótesis de que no hay diferencia en las localizaciones de las distribuciones poblacionales para los datos de la cantidad de bacterias del Ejemplo 15.4.

Solución Ya hemos indicado que la prueba U de Mann–Whitney y la prueba de suma de rangos de Wilcoxon son equivalentes, de modo que deberíamos llegar a las mismas conclusiones aquí al igual que en el Ejemplo 15.4. Recuerde que la hipótesis alternativa era que las distribuciones de cantidades de bacterias para los cultivos I y II diferían y que esto implicaba una prueba de las dos colas. Así, como la Tabla 8, Apéndice 3, da valores de $P(U \leq U_0)$ para tamaños muestrales especificados y valores de U_0, debemos duplicar el valor tabulado para hallar α. Suponga, como en el Ejemplo 15.4, que deseamos un valor de α cercano a .05. Verificando la Tabla 8 para ver $n_1 = n_2 = 4$, encontramos $P(U \leq 1) = .0286$. La región de rechazo apropiada para la prueba de dos colas es $U \leq 1$ o $U \geq n_1n_2 - 1 = 16 - 1 = 15$, para lo cual $\alpha = 2(.0286) = .0572$ o bien, redondeando a tres lugares decimales, $\alpha = .057$ (el mismo valor de α obtenido para el Ejemplo 15.4).

Para los datos de las bacterias, la suma de rango es $W = 12$. Entonces,

$$U = n_1n_2 + \frac{n_1(n_1 + 1)}{2} - W = (4)(4) + \frac{4(4 + 1)}{2} - 12 = 14 .$$

El valor calculado de U no cae en la región de rechazo. En consecuencia, no hay suficiente evidencia para demostrar una diferencia en las localizaciones de las distribuciones poblacionales de la cantidad de bacterias para los cultivos I y II. El valor p se da con $2P(U \geq 14) = 2P(U \leq 2) = 2(.0571) = .1142$. ■

EJEMPLO **15.6** Se realizó un experimento para comparar las resistencias de dos tipos de papeles de estraza, un papel estándar de un peso especificado y el otro del mismo papel estándar tratado con una sustancia química. Diez piezas de cada tipo de papel, seleccionadas aleatoriamente de un lote, produjeron las mediciones de resistencia que se muestran en la Tabla 15.5. Pruebe la hipótesis de que no hay diferencia en las distribuciones de resistencias para los dos tipos de papel, contra la hipótesis alternativa de que el papel tratado tiende a ser más fuerte.

Solución Ambas muestras son de tamaño 10, de modo que cada población (estándar o tratada) puede ser designada como población I. Hemos identificado las mediciones de papel estándar como

Tabla 15.5 Datos para el Ejemplo 15.6

Estándar, I	Tratado, II
1.21 (2)	1.49 (15)
1.43 (12)	1.37 (7.5)
1.35 (6)	1.67 (20)
1.51 (17)	1.50 (16)
1.39 (9)	1.31 (5)
1.17 (1)	1.29 (3.5)
1.48 (14)	1.52 (18)
1.42 (11)	1.37 (7.5)
1.29 (3.5)	1.44 (13)
1.40 (10)	1.53 (19)
Suma de rangos $W = 85.5$	

provenientes de la población I. En la Tabla 15.5 los rangos se muestran entre paréntesis junto a las $n_1 + n_2 = 10 + 10 = 20$ mediciones de resistencia y la suma de rangos W aparece debajo de la primera columna. Como deseamos detectar un desplazamiento en la distribución de la población I (estándar) a la izquierda de la distribución de la población II (tratada), rechazaremos la hipótesis nula de que no hay diferencia en las distribuciones poblacionales de resistencia cuando W es excesivamente pequeña. Debido a que esta situación ocurre cuando U es grande, realizaremos una prueba estadística de una cola y rechazaremos la hipótesis nula cuando $U \geq n_1 n_2 - U_0$.

Supóngase que escogemos un valor de α cercano a .05. Entonces podemos hallar U_0 al consultar la parte de la Tabla 8, Apéndice 3, correspondiente a $n_2 = 10$. La probabilidad $P(U \leq U_0)$ cercana a .05 es .0526 y corresponde a $U_0 = 28$. Por tanto, rechazaremos si $U \geq (10)(10) - 28 = 72$.

Al calcular U, tenemos

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W = (10)(10) + \frac{(10)(11)}{2} - 85.5 = 69.5.$$

Como se puede ver, U no es mayor que 72. Por tanto, no podemos rechazar la hipótesis nula. En el nivel de significancia $\alpha = .0526$, no hay suficiente evidencia para indicar que el papel de estraza tratado sea más fuerte que el estándar. El valor p está dado por $P(U \geq 69.5) = P(U \leq 30.5) = .0716$. ■

Una prueba simplificada para muestras grandes ($n_1 > 10$ y $n_2 > 10$) se puede obtener con el uso del ya conocido estadístico Z del Capítulo 10. Cuando las distribuciones poblacionales son idénticas, se puede demostrar que el estadístico U tiene los siguientes valor y varianza esperados:

$$E(U) = \frac{n_1 n_2}{2} \quad \text{y} \quad V(U) = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}.$$

También, cuando n_1 y n_2 son grandes,

$$Z = \frac{U - E(U)}{\sigma_U}$$

tiene aproximadamente una distribución normal estándar. Esta aproximación es adecuada cuando n_1 y n_2 son ambos mayores o iguales a 10. Así, para una prueba de dos colas con $\alpha = .05$, rechazaremos la hipótesis nula si $|z| \geq 1.96$.

El estadístico Z da la misma conclusión que la prueba U exacta para el Ejemplo 15.6:

$$z = \frac{69.5 - [(10)(10)/2]}{\sqrt{[(10)(10)(10+10+1)]/12}} = \frac{69.5 - 50}{\sqrt{2100/12}} = \frac{19.5}{\sqrt{175}}$$
$$= \frac{19.5}{13.23} = 1.47.$$

Para una prueba de una cola con $\alpha = .05$ ubicada en la cola superior de la distribución z, rechazaremos la hipótesis nula si $z > 1.645$. Se puede ver que $z = 1.47$ no cae en la región de rechazo y que esta prueba llega a la misma conclusión que la prueba U exacta del Ejemplo 15.6.

La prueba U de Mann–Whitney para muestras grandes $n_1 > 10$ y $n_2 > 10$

Hipótesis nula: H_0 : las distribuciones de frecuencia relativa para las poblaciones I y II son idénticas.

Hipótesis alternativa: (1) H_a : las distribuciones de frecuencia relativa de las dos poblaciones difieren en ubicación (una prueba de dos colas),

o bien, (2) la distribución de frecuencia relativa para la población I está desplazada a la derecha (o la izquierda) de la distribución de frecuencia relativa para la población II (una prueba de una cola).

Estadístico de prueba : $Z = \dfrac{U - (n_1 n_2/2)}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$.

Región de rechazo: rechazar H_0 si $z > z_{\alpha/2}$ o $z < -z_{\alpha/2}$ para una prueba de dos colas.

Para una prueba de una cola, ponga toda α en una cola de la distribución z. Para detectar un desplazamiento en la distribución de la población I a la derecha de la distribución de la población II, rechace H_0 cuando $z < -z_\alpha$. Para detectar un desplazamiento en la dirección contraria, rechace H_0 cuando $z > z_\alpha$. Los valores tabulados de z se dan en la Tabla 4, Apéndice 3.

Puede parecer que la prueba U de Mann–Whitney y la prueba de suma de rangos de Wilcoxon equivalente no son muy eficientes porque no parece que usen toda la información de la muestra. En realidad, estudios teóricos han demostrado que este no es el caso. Supongamos, por ejemplo, que todas las suposiciones para una prueba t de dos muestras se satisfacen cuando se prueba $H_0 : \mu_1 - \mu_2 = 0$ contra $H_a : \mu_1 - \mu_2 > 0$. Como la prueba t de dos muestras simplemente prueba para hallar una diferencia en localización (véase la Sección 15.2), podemos usar el estadístico U de Mann–Whitney para probar estas mismas hipótesis. Para una α y β dadas, el tamaño muestral total requerido para la prueba t es aproximadamente .95 multiplicado por el tamaño muestral total requerido para la U de Mann–Whitney. Así, el procedimiento no paramétrico es casi tan bueno como la prueba t para la situación en la que la prueba t es óptima. Para muchas distribuciones no normales, el procedimiento no paramétrico requiere menos observaciones que un procedimiento paramétrico correspondiente para producir los mismos valores de α y β.

Ejercicios

15.21 Encuentre los valores p asociados con cada una de las siguientes situaciones para probar H_0: las poblaciones I y II tienen la misma distribución.

a H_a : la distribución de la población I se desplaza a la derecha de la distribución de la población II; $n_1 = 4$, $n_2 = 7$, $W = 34$.

b H_a : la distribución de la población I se desplaza a la izquierda de la distribución de la población II; $n_1 = 5$, $n_2 = 9$, $W = 38$.

c H_a : las poblaciones I y II difieren en ubicación; $n_1 = 3$, $n_2 = 6$, $W = 23$.

15.22 En algunas pruebas de salud en ancianos, un nuevo medicamento ha mejorado la capacidad de memoria casi al nivel de los jóvenes. El medicamento pronto se probará en pacientes con el mal de Alzheimer, enfermedad cerebral mortal que finalmente destruye las mentes de aquellos que están afectados. De acuerdo con el doctor Gary Lynch de la Universidad de California en Irvine, el medicamento, llamado ampakino CX-516, acelera las señales entre células cerebrales y parece agudizar de manera significativa la memoria.[3] En una prueba preliminar hecha en estudiantes que apenas rebasan los 20 años y en hombres entre 65 y 70 años, los resultados fueron particularmente sorprendentes. Los siguientes datos son los números de sílabas incoherentes recordadas después de 5 minutos para diez hombres de 20 años o más y en diez hombres entre 65 y 70 años a quienes se les dio una dosis muy pequeña del ampakino CX-516. ¿Los datos aportan suficiente evidencia para concluir que hay una diferencia en el número de sílabas incoherentes recordadas por hombres de los dos grupos de edad cuando a los más viejos se les dio el ampakino CX-516? Determine el valor p asociado.

Grupo de edad	Número de sílabas recordadas									
20	11	7	6	8	6	9	2	10	3	6
65–70 (con ampakino CX-516)	1	9	6	8	7	8	5	7	10	3

15.23 Se probaron dos plásticos, cada uno de ellos producido por un proceso diferente, para hallar su resistencia máxima. Las mediciones de la siguiente tabla representan cargas de ruptura en unidades de 1000 libras por pulgada cuadrada. ¿Estos datos presentan evidencia de una diferencia entre las localizaciones de las distribuciones de resistencias máximas para los dos plásticos? Resuelva el problema usando la prueba U de Mann–Whitney con un nivel de significancia tan cercano como sea posible a $\alpha = 10$.

Plástico 1	Plástico 2
15.3	21.2
18.7	22.4
22.3	18.3
17.6	19.3
19.1	17.1
14.8	27.7

15.24 Los valores codificados para una medida de brillantez en papel (reflectividad ligera), elaborado por dos procesos diferentes, son como se muestra en la siguiente tabla para muestras de tamaño 9 obtenidas aleatoriamente de cada uno de los dos procesos. ¿Estos datos presentan suficiente evidencia para indicar una diferencia en la localización de las mediciones de brillantez para los dos procesos? Proporcione el nivel de significancia alcanzado.

3. *Fuente*: "Alzheimer's Test Set for New Memory Drug," *Press Enterprise* (Riverside, Calif.), 18 de noviembre de 1997, p. A-4

A	B
6.1	9.1
9.2	8.2
8.7	8.6
8.9	6.9
7.6	7.5
7.1	7.9
9.5	8.3
8.3	7.8
9.0	8.9

a Use la prueba U de Mann–Whitney.

b Use la prueba t de Student.

c Proporcione las hipótesis específicas nula y alternativa, junto con cualesquiera suposiciones, para las pruebas empleadas en los incisos a y b.

15.25 Se seleccionaron al azar 15 baterías experimentales de un lote de la planta piloto A y se seleccionaron 15 baterías estándar de la producción de la planta B. Las 30 baterías se colocaron simultáneamente bajo una carga eléctrica de la misma magnitud. La primera batería en fallar fue una A, la segunda una B, la tercera una B y así sucesivamente. La secuencia siguiente muestra el orden de falla para las 30 baterías:

A B B B A B A A B B B B A B A
B B B B A A B A A A B A A A A

Usando la teoría de muestras grandes para la prueba U determine si hay suficiente evidencia para permitir que el experimentador concluya que la vida útil para las baterías experimentales tiende a ser mayor que para las baterías estándar. Use $\alpha = .05$.

15.26 Consulte los Ejercicios 8.88 y 8.89. ¿Hay suficiente evidencia para indicar una diferencia en las poblaciones de mediciones de LC50 para DDT y Diazinon? ¿Cuál es el nivel de significancia alcanzado asociado con el estadístico U? ¿Qué se concluye cuando $\alpha = .10$?

15.27 A continuación aparece una tabla que indica frecuencias de aletazos[4] para muestras de dos especies de abejas euglosinas. Cuatro abejas de la especie *Euglossa mandibularis* Friese y seis de la especie *Euglossa imperialis* Cockerell se muestran en la siguiente tabla.

Frecuencias de aletazos	
E. mandibularis Friese	*E. imperialis* Cockerell
235	180
225	169
190	180
188	185
	178
	183

4. *Fuente*: T. M. Casey, M. L. May, and K. R. Morgan, "Flight Energetics of Euglossine Bees in Relation to Morphology and Wing Stroke Frequency," *Journal of Experimental Biology* 116 (1985).

a ¿Los datos presentan suficiente evidencia para indicar que las distribuciones de frecuencias de aletazos difieren para las dos especies? Use la prueba basada en el estadístico U de Mann–Whitney con α tan cercana a .10 como sea posible, pero sin rebasarlo.

b Dé el valor p aproximado asociado con la prueba.

15.28 El tratamiento de cáncer con quimioterapia emplea sustancias químicas que matan células cancerosas y normales. En algunos casos la toxicidad del medicamento contra el cáncer, es decir, su efecto en células normales, se puede reducir mediante la inyección simultánea de un segundo medicamento. Se realizó un estudio para determinar si la inyección de un medicamento particular era benéfica para reducir los efectos perjudiciales de un tratamiento de quimioterapia en el tiempo de supervivencia para ratas. Dos grupos de ratas seleccionados aleatoriamente, 12 ratas en cada grupo, se emplearon para el experimento. Ambos grupos, llamados A y B, recibieron el medicamento tóxico en una dosis lo suficientemente grande como para causar la muerte, pero el grupo B también recibió la antitoxina que estaba destinada a reducir el efecto tóxico de la quimioterapia en células normales. La prueba se concluyó al término de 20 días o 480 horas. Los tiempos de supervivencia para los dos grupos de ratas, a las 4 horas más cercanas, se muestran en la siguiente tabla. ¿Los datos arrojan evidencia suficiente para indicar que las ratas que recibieron la antitoxina tendían a vivir más después de la quimioterapia que las que no la recibieron? Use la prueba U de Mann–Whitney con un valor de α cercano a .05.

Sólo quimioterapia (A)	Quimioterapia más medicamento (B)
84	140
128	184
168	368
92	96
184	480
92	188
76	480
104	244
72	440
180	380
144	480
120	196

15.7 La prueba de Kruskal–Wallis para un diseño de un factor

En la Sección 13.3 presentamos un análisis de procedimiento de varianza (ANOVA) para comparar las medias de k poblaciones. La prueba F resultante estuvo basada en la suposición de que muestras aleatorias independientes se tomaron de poblaciones normales con varianzas iguales. Esto es, como dijimos en la Sección 15.2, estábamos interesados en probar si todas las poblaciones tenían la misma distribución contra la alternativa de que las poblaciones diferían en localización. Un elemento clave en el desarrollo del procedimiento era la cantidad identificada como la suma de cuadrados del tratamiento, SST. Como señalamos en el análisis de la Sección 13.3, cuanto mayor fuera el valor de la SST mayor sería el peso de la evidencia para rechazar la hipótesis nula de que las medias son todas iguales. En esta sección presentamos una técnica no paramétrica para probar si las poblaciones difieren en localización. Al igual que las otras técnicas no paramétricas analizadas en este capítulo, el procedimiento de Kruskal–Wallis no requiere suposiciones acerca de la forma real de las distribuciones de probabilidad.

Como en la Sección 13.3 suponemos que muestras aleatorias independientes han sido tomadas de k poblaciones que difieren sólo en localización, aunque no es necesario suponer que estas poblaciones poseen distribuciones normales. Para una generalización completa, permitimos que los tamaños muestrales sean desiguales y con n_i, para $i = 1, 2, \ldots, k$, representamos el tamaño de la muestra tomada de la i-ésima población. Igual que en el procedimiento de la Sección 15.5, combinamos todas las $n_1 + n_2 + \cdots + n_k = n$ observaciones y las clasificamos desde 1 (la más pequeña) hasta n (la más grande). Los empates se tratan como en las secciones previas. Esto es, si dos o más observaciones están empatadas para el mismo rango, entonces el promedio de los rangos que se hubieran asignado a estas observaciones se asigna a cada miembro del grupo empatado. Con R_i denotamos la suma de los rangos de las observaciones de la población i y con $\overline{R}_i = R_i/n_i$ denotamos el promedio correspondiente de los rangos. Si $\overline{R}$ es igual al promedio general de todos los rangos, consideramos el rango análogo SST, que se calcula usando los rangos en lugar de los valores reales de las mediciones:

$$V = \sum_{i=1}^{k} n_i(\overline{R}_i - \overline{R})^2.$$

Si la hipótesis nula es verdadera y las poblaciones no difieren en localización, esperaríamos que los valores $\overline{R}_i$ fueran aproximadamente iguales y el valor resultante de V fuera relativamente pequeño. Si la hipótesis alternativa es verdadera, esperaríamos que esto se refleje en las diferencias entre los valores de $\overline{R}_i$ que llevan a un valor grande para V. Observe que $\overline{R} =$ (suma de los primeros n enteros)$/n = [n(n + 1)/2]/n = (n + 1)/2$ y que

$$V = \sum_{i=1}^{k} n_i \left(\overline{R}_i - \frac{n+1}{2}\right)^2.$$

En lugar de concentrarse en V, Kruskal y Wallis (1952) consideraron el estadístico $H = 12V/[n(n + 1)]$, que se puede reescribir (véase el Ejercicio 15.35) como

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n + 1).$$

Como ya antes dijimos, la hipótesis nula de localizaciones iguales es rechazada a favor de la alternativa de que las poblaciones difieren en localización si el valor de H es grande. Así, la correspondiente prueba de nivel α pide el rechazo de la hipótesis nula a favor de la alternativa si $H > h(\alpha)$, donde $h(\alpha)$ es tal que, cuando H_0 es verdadera, $P[H > h(\alpha)] = \alpha$.

Si las distribuciones subyacentes son continuas y si no hay empates entre las n observaciones, la distribución nula de H puede hallarse (tediosamente) con el uso de los métodos del Capítulo 2. Podemos hallar la distribución de H para cualesquiera valores de k y $n_1, n_2, \ldots, n_k$ al calcular el valor de H para cada una de las $n!$ igualmente probables permutaciones de los rangos de las n observaciones (vea el Ejercicio 15.36). Estos cálculos se han efectuado ya y se han desarrollado tablas para algunos valores relativamente pequeños de k y para $n_1, n_2, \ldots, n_k$ [vea, por ejemplo, la Tabla A.12 de Hollander y Wolfe (1999)].

Kruskal y Wallis demostraron que si los n_i valores son "grandes", la distribución nula de H se puede aproximar con una distribución χ^2 con $k - 1$ grados de libertad (gl). Este cálculo se acepta generalmente como adecuado si cada uno de los n_i valores es mayor o igual a 5. Nuestros ejemplos y ejercicios son tales que esta aproximación de muestra grande es adecuado. Si usted desea usar el análisis de Kruskal–Wallis para conjuntos de datos más pequeños, donde esta aproximación para muestras grandes no es adecuada, consulte la obra de Hollander y Wolfe (1999) para obtener los valores críticos apropiados.

Resumimos el procedimiento de Kruskal-Wallis basada para muestrar grandes como sigue.

Prueba de Kruskal–Wallis basada en H para comparar k distribuciones poblacionales

Hipótesis nula: H_0: las k distribuciones poblacionales son idénticas.

Hipótesis alternativa: H_a: al menos dos de las distribuciones poblacionales difieren en localización.

Estadístico de prueba: $H = \{12/[n(n+1)]\}\sum_{i=1}^{k} R_i^2/n_i - 3(n+1)$, donde

n_i = número de mediciones en la muestra de la población i,

R_i = suma de rangos para la muestra i, donde el rango de cada medida se calcula de acuerdo con su tamaño relativo en el conjunto general de $n = n_1 + n_2 + \cdots + n_k$ observaciones formadas al combinar los datos de todas las k muestras.

Región de rechazo: rechazar H_0 si $H > \chi^2_\alpha$ con $(k-1)$ grados de libertad.

Suposiciones: las k muestras se sacan al azar y en forma independiente. Hay cinco o más mediciones en cada muestra.

EJEMPLO **15.7** Un ingeniero de control de calidad ha seleccionado muestras independientes de la producción de tres líneas de ensamble en una planta de componentes electrónicos. Para cada línea, la producción de diez horas seleccionadas al azar se examinó en busca de defectos. ¿Los datos de la Tabla 15.6 dan evidencia de que las distribuciones de probabilidad del número de defectos por hora de producción difieren en localización para al menos dos de las líneas? Use $\alpha = .05$. También proporcione el valor p asociado con la prueba.

Solución En este caso, $n_1 = 10 = n_2 = n_3$ y $n = 30$. Así,

$$H = \frac{12}{30(31)}\left[\frac{(120)^2}{10} + \frac{(210.5)^2}{10} + \frac{(134.5)^2}{10}\right] - 3(31) = 6.097.$$

Tabla 15.6 Datos para el Ejemplo 15.7

Línea 1		Línea 2		Línea 3	
Defectos	Rango	Defectos	Rango	Defectos	Rango
6	5	34	25	13	9.5
38	27	28	19	35	26
3	2	42	30	19	15
17	13	13	9.5	4	3
11	8	40	29	29	20
30	21	31	22	0	1
15	11	9	7	7	6
16	12	32	23	33	24
25	17	39	28	18	14
5	4	27	18	24	16
	$R_1 = 120$		$R_2 = 210.5$		$R_3 = 134.5$

Debido a que todos los valores n_i son mayores o iguales a 5, podemos usar la aproximación para la distribución nula de H y rechazar la hipótesis nula de iguales localizaciones si $H > \chi^2_\alpha$ con base en $k - 1 = 2$ grados de libertad. Consultamos la Tabla 6, Apéndice 3, para determinar que $\chi^2_{.05} = 5.99147$. Por tanto, rechazamos la hipótesis nula en el nivel $\alpha = .05$ y concluimos que al menos una de las tres líneas tiende a producir un número mayor de piezas defectuosas que las otras.

De acuerdo con la Tabla 6, Apéndice 3, el valor de $H = 6.097$ lleva al rechazo de la hipótesis nula si $\alpha = .05$ pero no si $\alpha = .025$. Entonces, $.025 <$ valor $p < .05$. La aplicación *Chi-Square Probability and Quantiles* se puede usar para establecer que el valor p aproximado es igual a $P(\chi^2 > 6.097) = .0474$. ■

Se puede demostrar que si deseamos comparar sólo $k = 2$ poblaciones, la prueba de Kruskal–Wallis es equivalente a la prueba bilateral de Wilcoxon de suma de rangos presentada en la Sección 15.5. Si se obtienen datos de un diseño de un factor que comprenda $k > 2$ poblaciones pero deseamos comparar un par de poblaciones en particular, la prueba de suma de rangos de Wilcoxon (o la prueba equivalente U de Mann–Whitney de la Sección 15.6) se puede usar para este fin. Observe que el análisis basado en el estadístico H de Kruskal–Wallis no requiere el conocimiento de los valores reales de las observaciones. Sólo es necesario conocer los rangos de las observaciones para completar el análisis. El Ejercicio 15.32 ilustra el uso del análisis de Kruskal–Wallis para tal caso.

Ejercicios

15.29 La siguiente tabla contiene información sobre la longitud de las hojas de plantas de la misma especie en cada uno de cuatro lugares pantanosos no urbanizados. En cada uno de éstos, se seleccionaron aleatoriamente seis plantas. Se seleccionaron al azar diez hojas de cada una de las plantas y la media de las diez mediciones (en centímetros) se registró para cada planta de cada uno de los lugares. Use la prueba H de Kruskal–Wallis para determinar si hay suficiente evidencia para afirmar que la distribución de las longitudes medias de las hojas difiere en localizaciones para al menos dos de los sitios. Use $\alpha = .05$. Establezca el límite o encuentre el valor p aproximado.

Localización	Longitud media de la hoja (cm)					
1	5.7	6.3	6.1	6.0	5.8	6.2
2	6.2	5.3	5.7	6.0	5.2	5.5
3	5.4	5.0	6.0	5.6	4.0	5.2
4	3.7	3.2	3.9	4.0	3.5	3.6

15.30 Una compañía piensa promover un nuevo producto mediante el uso de una de tres campañas publicitarias. Para investigar la magnitud del reconocimiento del producto que resulte de las campañas, se seleccionaron 15 áreas de mercado y a cada 5 se les asignó aleatoriamente una campaña. Al final de las campañas se seleccionaron muestras aleatorias de 400 adultos en cada área y las proporciones de quienes indicaron familiaridad con el producto aparecen en la siguiente tabla.

Campaña		
1	2	3
.33	.28	.21
.29	.41	.30
.21	.34	.26
.32	.39	.33
.25	.27	.31

a ¿Qué tipo de diseño experimental se utilizó?

b ¿Hay suficiente evidencia para indicar la diferencia en ubicaciones de las distribuciones de calificaciones de reconocimiento de producto para las tres campañas? Establezca el límite o dé el valor de p aproximado.

c Las campañas 2 y 3 fueron, respectivamente, las más y menos costosas. ¿Hay suficiente evidencia para indicar que la campaña 2 es más exitosa que la campaña 3? Pruebe usando el procedimiento U de Mann–Whitney. Proporcione el valor p asociado.

15.31 Tres diferentes marcas de tubos magnetrón (componentes clave en hornos de microondas) fueron sometidas a pruebas de esfuerzo, registrándose el número de horas que cada uno de ellos operó sin necesidad de reparación (véase la tabla siguiente). Aunque estos tiempos no representan tiempos de vida útil típicos, indican lo bien que estos tubos pueden resistir esfuerzos extremos.

Marca A	Marca B	Marca C
36	49	71
48	33	31
5	60	140
67	2	59
53	55	42

a Use la prueba F para un diseño de un factor (Capítulo 13) para probar la hipótesis de que la duración media con esfuerzo es igual para las tres marcas. Use $\alpha = .05$. ¿Qué suposiciones son necesarias para la validez de este procedimiento? ¿Hay alguna razón para dudar de estas suposiciones?

b Use la prueba Kruskal–Wallis para determinar si existe evidencia para concluir que las marcas de tubos magnetrón tienden a diferir en duración con esfuerzo. Pruebe usando $\alpha = .05$.

15.32 Se realizó un experimento para comparar el tiempo que una persona tarda en recuperarse de cada uno de los tres tipos de gripe, Victoria A, Texas y Rusa. Veintiún personas fueron seleccionadas aleatoriamente de un grupo de voluntarios y divididas en tres grupos de 7 cada uno. A cada grupo se le asignó al azar una variedad del virus y la gripe fue inducida en las personas, todas las cuales recibieron atención médica en condiciones idénticas y se registró el tiempo de recuperación (en días). Los rangos de los resultados aparecen en la siguiente tabla.

Victoria A	Texas	Rusa
20	14.5	9
6.5	16.5	1
21	4.5	9
16.5	2.5	4.5
12	14.5	6.5
18.5	12	2.5
9	18.5	12

a ¿Los datos aportan suficiente evidencia para indicar que los tiempos de recuperación para uno (o más) tipo(s) de gripe tiende(n) a ser más largos que para los otros tipos? Proporcione el valor p asociado.

b ¿Los datos aportan suficiente evidencia para indicar una diferencia en las localizaciones de las distribuciones de tiempos de recuperación para los tipos Victoria A y Rusa? Proporcione el valor p asociado.

15.33 La EPA desea determinar si los cambios de temperatura en el agua de mar causados por una planta nuclear para generar energía eléctrica, tendrán un efecto importante en la fauna de la región. Recientemente, especímenes recién nacidos de ciertos peces se han dividido aleatoriamente en cuatro grupos que se colocan en ambientes oceánicos simulados y separados que son idénticos en todo excepto en la temperatura del agua. Seis meses después se pesan los especímenes y los resultados (en onzas) se dan en la siguiente tabla. ¿Los datos arrojan suficiente evidencia para indicar que una (o más) de las temperaturas tiende a producir aumentos de peso mayores que las otras temperaturas? Pruebe usando $\alpha = .10$.

Pesos de especímenes			
38°F	42°F	46°F	50°F
22	15	14	17
24	21	28	18
16	26	21	13
18	16	19	20
19	25	24	21
	17	23	

15.34 Cada año los gorgojos ocasionan pérdidas de millones de dólares en las cosechas de algodón. Se aplican tres sustancias químicas diseñadas para controlar las poblaciones de gorgojos. Después de 3 meses, diez lotes de igual tamaño se seleccionan aleatoriamente dentro de cada campo y, para cada uno, se registra el porcentaje de plantas de algodón dañadas por los gorgojos. ¿Los datos de la siguiente tabla aportan suficiente evidencia para indicar una diferencia en la localización entre las distribuciones de porcentajes de daños correspondientes a los tres tratamientos? Establezca límites para el valor p asociado.

Sustancia química A	Sustancia química B	Sustancia química C
10.8	22.3	9.8
15.6	19.5	12.3
19.2	18.6	16.2
17.9	24.3	14.1
18.3	19.9	15.3
9.8	20.4	10.8
16.7	23.6	12.2
19.0	21.2	17.3
20.3	19.8	15.1
19.4	22.6	11.3

15.35 El estadístico de Kruskal–Wallis es

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left(\overline{R}_i - \frac{n+1}{2} \right)^2 .$$

Efectúe la operación indicada de elevar al cuadrado cada término de la suma y, además, sume los valores resultantes para demostrar que

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1).$$

[*Sugerencia*: recuerde que $\overline{R}_i = R_i/n_i$ y que $\sum_{i=1}^{k} R_i =$ suma de los primeros n enteros $= n(n+1)/2$.]

15.36 Suponiendo que no haya empates, obtenga la distribución nula exacta del estadístico H de Kruskal–Wallis para el caso $k = 3$, $n_1 = n_2 = n_3 = 2$. [Debido a que los tamaños muestrales son todos iguales, si los rangos 1 y 2 se asignan al tratamiento 1, los rangos 3 y 4 se asignan al tratamiento 2 y los rangos 5 y 6 se asignan al tratamiento 3, el valor de H es exactamente el mismo que si los rangos 3 y 4 se asignaran al tratamiento 1, los rangos 5 y 6 se asignaran al tratamiento 2 y los rangos 1 y 2 se asignaran al tratamiento 3. Esto es, para cualquier conjunto particular de rangos, podemos intercambiar los papeles de las k poblaciones y obtener los mismos valores del estadístico H. Entonces, el número de casos que debemos considerar puede ser reducido por un factor de $1/k!$. En consecuencia, H debe ser evaluado sólo para $(6!/[2! \cdot 2! \cdot 2!])/3! = 15$ arreglos distintos de rangos.]

15.8 La prueba de Friedman para diseños de bloques aleatorizados

En la Sección 12.4 estudiamos los méritos de un diseño de bloques aleatorizado para un experimento que compara el rendimiento de varios tratamientos. Suponemos que b bloques se emplean en el experimento, que está diseñado para comparar las localizaciones de las distribuciones de las respuestas correspondientes a cada uno de k tratamientos. El ANOVA, que estudiamos en la Sección 13.9, estuvo basado en las suposiciones de que las observaciones en cada combinación de bloque-tratamiento estuvieron distribuidas normalmente con varianzas iguales. Al igual que en el caso del diseño de un factor, la SST fue la cantidad clave en el análisis.

La prueba de Friedman, inventada por Milton Friedman (1937), economista ganador del premio Nobel, está diseñada para probar la hipótesis nula de que las distribuciones de probabilidad de los k tratamientos son idénticas contra la alternativa de que al menos dos de las distribuciones difieren en localización. La prueba está basada en un estadístico equivalente en rangos a la SST para el diseño de bloques aleatorizado (vea la Sección 13.9) y se calcula en la siguiente forma. Después de obtener los datos de un diseño de bloques aleatorizado, *dentro de cada bloque* los valores observados de las respuestas a cada uno de los k tratamientos se clasifican de 1 (el más pequeño del bloque) a k (el más grande del bloque). Si dos o más observaciones *del mismo bloque* están empatadas para el mismo rango, entonces el promedio de los rangos que se hubieran asignado a estas observaciones se asigna a cada miembro del grupo empatado. No obstante, los empates necesitan resolverse de este modo sólo si se presentan dentro del mismo bloque.

Con R_i denote la suma de los rangos de las observaciones correspondientes al tratamiento i y con $\overline{R}_i = R_i/b$ denote el promedio correspondiente de los rangos (recuerde que, en un diseño de bloques aleatorizado, cada tratamiento se aplica exactamente una vez en cada bloque, resultando en un total de b observaciones por tratamiento y por tanto en un total de bk observaciones). Debido a que los rangos de 1 a k están asignados dentro de cada bloque, la suma de los rangos asignados en cada bloque es $1 + 2 + \cdots + k = k(k + 1)/2$. Así, la suma de todos los rangos asignados en el análisis es $bk(k + 1)/2$. Si $\overline{R}$ denota el promedio general de los rangos de todas las observaciones bk, se deduce que $\overline{R} = (k + 1)/2$. Considere el equivalente en rangos de la SST para un diseño de bloques aleatorizado dado por

$$W = b\sum_{i=1}^{k}(\overline{R}_i - \overline{R})^2.$$

Si la hipótesis nula es verdadera y las distribuciones de probabilidad de las respuestas del tratamiento no difieren en localización, esperamos que los valores $\overline{R}_i$ sean aproximadamente iguales y el valor resultante para W sea pequeño. Si la hipótesis alternativa fuera verdadera, esperaríamos que esto llevara a diferencias entre los valores $\overline{R}_i$ y grandes valores de W correspondientes. En lugar de W, Friedman consideró el estadístico $F_r = 12W/[k(k+1)]$, que se puede reescribir (véase el Ejercicio 15.44) como

$$F_r = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1).$$

Como vimos antes, la hipótesis nula de localizaciones iguales es rechazada a favor de la alternativa de que las distribuciones de tratamiento difieren en localización si el valor de F_r es grande. Esto es, la prueba de nivel α correspondiente rechaza la hipótesis nula a favor de la alternativa si $F_r > f_r(\alpha)$, donde $f_r(\alpha)$ es tal que, cuando H_0 es verdadera, $P[F_r > f_r(\alpha)] = \alpha$.

Si no hay empates entre las observaciones dentro de los bloques, la distribución nula de F_r puede (tediosamente) hallarse con el uso de los métodos del Capítulo 2. Para cualesquiera valores de b y k, la distribución de F_r se halla como sigue. Si la hipótesis nula es verdadera, entonces cada una de las $k!$ permutaciones de los rangos $1, 2, \ldots, k$ dentro de cada bloque es igualmente probable. Además, como suponemos que las observaciones en bloques diferentes son mutuamente independientes, se deduce que cada una de las $(k!)^b$ posibles combinaciones de los b conjuntos de permutaciones para los rangos dentro del bloque son igualmente probables cuando H_0 es verdadera. En consecuencia, podemos evaluar el valor de F_r para cada caso posible y por tanto dar la distribución nula de F_r (véase el Ejercicio 15.45). Los valores seleccionados de $f_r(\alpha)$ para varias opciones de k y b se muestran en la Tabla A.22 de Hollander y Wolfe (1999). Al igual que los otros procedimientos no paramétricos estudiados en este capítulo, la ventaja real de este procedimiento es que se puede usar cualquiera que sea la forma de las distribuciones reales de las poblaciones correspondientes a los tratamientos.

Como fue el caso con el estadístico de Kruskal–Wallis, la distribución nula del estadístico F_r de Friedman se puede calcular con una distribución χ^2 con $k-1$ grados de libertad mientras b sea "grande." La evidencia empírica indica que la aproximación es adecuada si b (el número de bloques) o k (el número de tratamientos) es mayor que 5. De nuevo, nuestros ejemplos y ejercicios resuelven situaciones en las que esta aproximación con muestras grandes es adecuada. Si usted necesita poner en práctica un análisis de Friedman para muestras pequeñas, consulte la obra de Hollander y Wolfe (1999) para obtener valores críticos apropiados.

Prueba de Friedman basada en F_r para un diseño de bloques aleatorizado

Hipótesis nula: H_0: las distribuciones de probabilidad para los k tratamientos son idénticas.

Hipótesis alternativa: H_a: al menos dos de las distribuciones difieren en localización.

Estadístico de prueba: $F_r = \{12/[bk(k+1)]\}\sum_{i=1}^{k} R_i^2 - 3b(k+1)$, donde

b = número de bloques,
k = número de tratamientos,
R_i = suma de los rangos para el i-ésimo tratamiento, donde el rango de cada medición se calcula con respecto a su tamaño dentro de su propio bloque.

Región de rechazo: $F_r > \chi^2_\alpha$ con $(k-1)$ grados de libertad.

Suposiciones: los tratamientos se asignan al azar a unidades experimentales dentro de los bloques. Ya sea que el número de bloques (b) o el número de tratamientos (k) excedan de 5.

EJEMPLO **15.8** Un experimento para comparar los tiempos de terminación para tres trabajos técnicos se realizó de la siguiente manera. Debido a que los tiempos de terminación pueden variar en forma considerable de una persona a otra, a cada uno de los seis técnicos se le pidió ejecutar los tres trabajos, mismos que fueron presentados a cada técnico en orden aleatorio con retraso apropiado entre los trabajos. ¿Los datos de la Tabla 15.7 presentan suficiente evidencia para indicar que las distribuciones de tiempos de terminación para los tres trabajos difieren en localización? Use $\alpha = .05$. Establezca límites para el valor p asociado.

Tabla 15.7 Tiempos de terminación para tres trabajos

Técnico	Trabajo A	Rango	Trabajo B	Rango	Trabajo C	Rango
1	1.21	1	1.56	3	1.48	2
2	1.63	1.5	2.01	3	1.63	1.5
3	1.42	1	1.70	2	2.06	3
4	1.16	1	1.27	2.5	1.27	2.5
5	2.43	2	2.64	3	1.98	1
6	1.94	1	2.81	3	2.44	2
		$R_1 = 7.5$		$R_2 = 16.5$		$R_3 = 12$

Solución El experimento fue ejecutado de acuerdo con un diseño de bloques aleatorizado con técnicos desempeñando el papel de bloques. En este caso, $k = 3$ tratamientos se compararon usando $b = 6$ bloques. Como el número de bloques excede de 5, podemos usar el análisis de Friedman y comparar el valor de F_r contra χ^2_α, con base en $k - 1 = 2$ grados de libertad. Si se consulta la Tabla 6, Apéndice 3, encontramos $\chi^2_{.05} = 5.99147$. Para los datos dados en la Tabla 15.7,

$$F_r = \frac{12}{6(3)(4)}[(7.5)^2 + (16.5)^2 + (12)^2] - 3(6)(4) = 6.75.$$

Como $F_r = 6.75$, que excede de 5.99147, concluimos con un nivel de $\alpha = .05$ que los tiempos de terminación de al menos dos de los tres trabajos poseen distribuciones de probabilidad que difieren en localización.

Como $F_r = 6.75$ es el valor observado de un estadístico que tiene aproximadamente una distribución χ^2 con 2 grados de libertad, se deduce que (aproximadamente) $.025 <$ valor $p < .05$. La aplicación *Chi-Square Probability and Quantiles* aplica para establecer que el valor p aproximado $= P(\chi^2 > 6.75) = .0342$. ■

En algunas situaciones podría ser fácil clasificar las respuestas dentro de cada bloque, pero mucho más difícil asignar un valor numérico significativo a la respuesta para cada tratamiento en los bloques. Un ejemplo que ilustra esta situación aparece en el Ejercicio 15.42.

Se puede ver (véase el Ejercicio 15.43) que, si deseamos comparar sólo $k = 2$ tratamientos usando un diseño de bloques aleatorizado (para que los bloques sean de tamaño 2), el estadístico de Friedman es el cuadrado del estadístico estandarizado de signo (esto es, el cuadrado del estadístico Z dado en la Sección 15.3). Entonces, para $k = 2$, el análisis de Friedman es equivalente a una prueba de signo de dos colas.

Ejercicios

15.37 En un estudio del sabor de antibióticos para niños, Doreen Matsui y sus colegas emplearon una muestra voluntaria de niños sanos para evaluar sus reacciones al sabor de cuatro antibióticos.[5] Las respuestas de los niños se midieron en una escala analógica visual de 10 centímetros que incorporaba el uso de caras, de triste (calificación baja) a alegre (calificación alta). Las calificaciones mínima y máxima fueron, respectivamente, 0 y 10. Los datos de la siguiente tabla (simulados a partir de los resultados dados en el informe de Matsui) se obtuvieron cuando a cada uno de cinco niños se les preguntó la calificación del gusto de los cuatro antibióticos.

	Antibiótico			
Niño	I	II	III	IV
1	4.8	2.2	6.8	6.2
2	8.1	9.2	6.6	9.6
3	5.0	2.6	3.6	6.5
4	7.9	9.4	5.3	8.5
5	3.9	7.4	2.1	2.0

a ¿Hay suficiente evidencia para concluir que hay diferencias en el sabor percibido de los diferentes antibióticos? Establezca un límite o encuentre el valor p aproximado.

b ¿Qué se concluiría en el nivel de significancia $\alpha = .05$?

c ¿Por qué Matsui hizo que cada niño ordenara los cuatro antibióticos en lugar de usar 20 niños diferentes, seleccionando aleatoriamente 5 para que recibieran sólo el antibiótico I, otros 5 para que recibieran sólo el antibiótico II, 5 de los restantes para que recibieran sólo el antibiótico III, con los 5 restantes recibiendo sólo el antibiótico IV?

15.38 Se efectuó un experimento para evaluar si se acumulan metales pesados en plantas que crecen en suelos mejorados con lodos y si hay una acumulación asociada de esos metales en pulgones que se alimentan de esas plantas.[6] Los datos de la siguiente tabla son concentraciones de cadmio (en microgramos/kilogramo) en plantas que crecen con seis diferentes magnitudes de aplicación de lodos para tres diferentes cosechas. Las magnitudes de aplicación son los tratamientos y las tres cosechas representan bloques de tiempo.

	Cosecha		
Rapidez	1	2	3
Control	162.1	153.7	200.4
1	199.8	199.6	278.2
2	220.0	210.7	294.8
3	194.4	179.0	341.1
4	204.3	203.7	330.2
5	218.9	236.1	344.2

5. *Fuente*: D. Matsui y otros, "Assessment of the Palatability of β-Lactamase-Resistant Antibiotics in Children," *Archives of Pediatric Adolescent Medicine* 151(1997): 559-601.

6. *Fuente*: G. Merrington, L. Winder, and I. Green, "The Uptake of Cadmium and Zinc by the Birdcherry Oat Aphid *Rhopalosiphum Padi* (*Homoptera: Aphididae*) Feeding on Wheat Grown on Sewage Sludge Amended Agricultural Soil," *Environmental Pollution* 96(1) (1997): 111-114.

a ¿Hay suficiente evidencia para indicar una diferencia en la acumulación de cadmio en plantas que crecen en lotes sometidos a diferentes niveles de aplicación de lodos? Establezca un límite o determine el valor p aproximado.

b ¿Qué se concluiría en el nivel de significancia $\alpha = .01$?

15.39 La corrosión de metales es un problema en numerosos aparatos mecánicos. Tres selladores que se usan para ayudar a retardar la corrosión de metales se probaron para ver si hubo algunas diferencias entre ellos. Muestras de diez lingotes diferentes de la misma composición de metal se trataron con cada uno de los tres selladores y la cantidad de corrosión se midió después de exponerlos a las mismas condiciones ambientales durante un mes. Los datos se muestran en la siguiente tabla. ¿Hay alguna evidencia de una diferencia en la capacidad de los selladores para evitar la corrosión? Pruebe usando $\alpha = .05$.

	Sellador		
Lingote	I	II	III
1	4.6	4.2	4.9
2	7.2	6.4	7.0
3	3.4	3.5	3.4
4	6.2	5.3	5.9
5	8.4	6.8	7.8
6	5.6	4.8	5.7
7	3.7	3.7	4.1
8	6.1	6.2	6.4
9	4.9	4.1	4.2
10	5.2	5.0	5.1

15.40 Un grave problema relacionado con las sequías para los agricultores es la diseminación de la aflatoxina, sustancia altamente tóxica generada por hongos y que contamina el maíz en los campos. En niveles más altos de contaminación, la aflatoxina es riesgosa para los animales y posiblemente para la salud humana. (Oficiales de la FDA han establecido un límite máximo de 20 partes por mil millones de aflatoxina como seguro para ventas interestatales.) Tres sustancias que se aplican con aspersor, A, B y C, se han inventado para controlar la aflatoxina en los campos de maíz. Para determinar si existen diferencias entre estas sustancias, diez mazorcas de maíz se escogen al azar de un campo de maíz contaminado y cada una se corta en tres partes de igual tamaño. Las sustancias se asignan aleatoriamente a las partes de cada mazorca de maíz, formando así un diseño de bloques aleatorizado. La siguiente tabla muestra la cantidad (en partes por mil millones) de aflatoxina presente en las muestras de maíz después de rociarlas. Use la prueba de Friedman basada en F_r para determinar si hay diferencias entre las sustancias para el control de la aflatoxina. Establezca límites aproximados para el valor p.

	Sustancia				Sustancia		
Mazorca	A	B	C	Mazorca	A	B	C
1	21	23	15	6	5	12	6
2	29	30	21	7	18	18	12
3	16	19	18	8	26	32	21
4	20	19	18	9	17	20	9
5	13	10	14	10	4	10	2

15.41 Se realizó un estudio para comparar las preferencias de ocho "expertos en escuchar" respecto a 15 modelos (con precios de lista aproximadamente iguales) de un componente particular de un sistema estéreo. Se hizo todo esfuerzo posible por asegurar que las diferencias percibidas por los escuchas se debieran al componente de interés y no a otra causa (todos los otros componentes del sistema eran idénticos, se usó el mismo tipo de música, la música se reprodujo en el mismo cuarto, etc.). Entonces, los resultados de la prueba de escucha reflejan las preferencias de audio de los jueces y no juicios respecto a la calidad, confiabilidad u otras variables. Además, los resultados están relacionados sólo con los modelos de los componentes usados en el estudio y no a ningún otro modelo que pueda ser ofrecido por los diversos fabricantes. Los datos de la siguiente tabla muestran los resultados de las pruebas de escucha. Los modelos se describen simplemente como A, B, . . . , O. Bajo cada encabezado de columna están los números de los jueces que clasificaron cada una de las marcas de componente de 1 (rango más bajo) a 15 (rango más alto).

	Rango														
Modelo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
B	0	0	0	1	0	2	1	1	1	0	0	0	0	2	0
C	0	1	1	1	4	0	0	1	0	0	0	0	0	0	0
D	1	0	1	1	0	1	0	0	1	0	1	0	1	1	0
E	0	2	1	3	0	2	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	1	2	2	3	0	0	0
G	0	0	0	0	0	0	0	0	0	1	0	2	4	1	0
H	1	2	1	1	0	0	2	1	0	0	0	0	0	0	0
I	3	2	1	0	0	0	0	0	0	1	0	1	0	0	0
J	0	0	1	0	2	0	2	0	0	0	2	0	1	0	0
K	0	0	0	0	0	0	1	1	0	2	1	1	1	1	0
L	0	0	0	0	0	0	1	1	4	0	1	0	1	0	0
M	1	1	2	1	1	2	0	0	0	0	0	0	0	0	0
N	2	0	0	0	0	0	1	1	0	0	0	1	0	3	0
O	0	0	0	0	1	1	0	2	1	2	1	0	0	0	0

a Use el procedimiento de Friedman para probar si las distribuciones de las calificaciones de preferencia difieren en localización para los 15 modelos del componente. Establezca limites para el nivel de significancia alcanzado. ¿Qué se concluiría en el nivel de significancia de $\alpha = .01$? [*Sugerencia*: la suma de los rangos asociada con el componente del modelo O es $5 + 6 + 8 + 8 + 9 + 10 + 10 + 11 = 67$; otras sumas de rango se pueden calcular de una manera análoga.]

b Si, antes de ejecutar el experimento deseamos comparar los componentes de los modelos G y H, esta comparación podría hacerse usando la prueba de signos presentada en la Sección 15.3. Usando la información recién dada, podemos determinar que el modelo G fue preferido al modelo H por los ocho jueces. Explique por qué. Proporcione el nivel de significancia alcanzado si la prueba de signos se usó para comparar componentes de los modelos G y H.

c Explique por qué no hay suficiente información para usar la prueba de signos en una comparación sólo de los modelos H y M.

15.42 Se realiza un experimento para investigar el efecto tóxico de tres sustancias químicas, A, B y C, en la piel de ratas. Tres cuadros adyacentes de $\frac{1}{2}$ pulgada se marcan en los lomos de ocho ratas y cada una de las tres sustancias se aplica a cada rata. Los cuadros de piel de cada rata se clasifican de acuerdo con la severidad de la irritación ($1 =$ menos severa, $3 =$ severidad máxima). Los datos resultantes se dan en la siguiente tabla. ¿Hay suficiente evidencia para apoyar la hipótesis de investigación de que las distribuciones de probabilidad de las calificaciones de irritación de la piel, correspondientes a las tres sustancias químicas, difieren en localización? Use $\alpha = .01$. (*Nota*: clasificar la severidad de las reacciones a las sustancias químicas para cada rata es probablemente mucho más significativo que asignar una "calificación de irritación" a cada parte de piel.)

	Sustancia química		
Rata	A	B	C
1	3	2	1
2	3	2	1
3	2	3	1
4	1	3	2
5	1	2	3
6	1	3	2
7	2	3	1
8	2	1	3

15.43 Considere el estadístico F_r de Friedman, cuando $k = 2$ y b = (número de bloques) = n. Entonces, $F_r = (2/n)(R_1^2 + R_2^2) - 9n$. Sea M el número de bloques (pares) en los que el tratamiento uno tiene la clasificación 1. Si no hay empates, entonces el tratamiento 1 tiene clasificación 2 en los restantes $n - M$ pares. Así, $R_1 = M + 2(n - M) = 2n - M$. De manera análoga, $R_2 = n + M$. Sustituya estos valores en la expresión anterior para F_r y demuestre que el valor resultante es $4(M - .5n)^2/n$. Compare este resultado con el cuadrado del estadístico Z de la Sección 15.3. Este procedimiento demuestra que $F_r = Z^2$.

15.44 Considere el estadístico de Friedman

$$F_r = \frac{12b}{k(k+1)} \sum_{i=1}^{k} (\overline{R}_i - \overline{R})^2$$

Eleve al cuadrado cada término de la suma y demuestre que una forma alternativa de F_r es

$$F_r = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1).$$

[*Sugerencia:* recuerde que $\overline{R}_i = R_i/b$, $\overline{R} = (k+1)/2$ y observe que $\sum_{i=1}^{k} R_i$ = suma de todos los rangos = $bk(k+1)/2$].

15.45 Si no hay empates y $b = 2$, $k = 3$, deduzca la distribución nula exacta de F_r.

15.9 Prueba de corridas de ensayo: una prueba de aleatoriedad

Considere un proceso de producción en el que piezas manufacturadas salen en secuencia y cada una es clasificada como defectuosa (D) o no defectuosa (N). Hemos estudiado cómo podríamos comparar la fracción de piezas defectuosas para dos intervalos iguales de tiempo, con el uso de una prueba Z (Capítulo 10) y ampliado esta prueba para probar la hipótesis de p constante pasa dos o más intervalos de tiempo usando la prueba χ^2 del Capítulo 14. Los propósitos de estas pruebas eran detectar un cambio o tendencia en la fracción de piezas defectuosas, p. La evidencia para indicar una fracción creciente de piezas defectuosas podría indicar la necesidad de un estudio del proceso para localizar la fuente del problema. Un valor decreciente podría sugerir que un programa de control de calidad del proceso estaba teniendo un efecto benéfico para reducir la fracción de piezas defectuosas.

Las tendencias en la fracción de piezas defectuosas (u otras medidas de calidad) no son la única indicación de falta de control de un proceso. Un proceso podría estar causando corridas periódicas de piezas defectuosas aunque la fracción promedio de piezas defectuosas permanezca constante, para todos los fines prácticos, en periodos largos. Por ejemplo, los focos de proyectores son fabricados en una máquina giratoria con un número fijo de posiciones para focos. Uno de éstos se coloca en la máquina en una posición determinada, se extrae el aire, se bombean gases en el interior del foco y la base de vidrio se sella con fuego. Si una máquina contiene 20 posiciones y varias posiciones adyacentes son defectuosas (quizá debido a demasiado calor empleado en el proceso de sellado), oleadas de focos defectuosos saldrán del proceso en forma periódica. Las pruebas que comparan la fracción del proceso de piezas defectuosas producidas durante intervalos iguales de tiempo no detectarán esta dificultad periódica del proceso. Esta periodicidad, indicada por las corridas de piezas defectuosas, indica que no hay aleatoriedad en la aparición de piezas defectuosas a lo largo del tiempo y puede ser detectada por una *prueba de aleatoriedad.* La prueba estadística que presentamos, conocida como *prueba de corridas de ensayo*, es estudiada en detalle por Wald y Wolfowitz (1940). Seguirán otras aplicaciones prácticas de la prueba de corridas de ensayo.

Como su nombre lo indica, las pruebas de corridas de ensayo se usan para estudiar una secuencia de eventos en la que cada elemento de la secuencia puede tomar uno de dos resultados, no defectuoso (S) o defectuoso (F). Si consideramos la sucesión de piezas que salen de un proceso de manufactura como defectuosas (F) o no defectuosas (S), la observación de veinte piezas podría dar

$$\begin{array}{cccccccccc} S & S & S & S & S & F & F & S & S & S \\ F & F & F & S & S & S & S & S & S & S. \end{array}$$

Observamos las agrupaciones de piezas defectuosas y no defectuosas y nos preguntamos si esta agrupación implica no aleatoriedad y, en consecuencia, falta de control de proceso.

DEFINICIÓN **15.1** Una *corrida de ensayo* es una subsecuencia máxima de elementos semejantes.

Por ejemplo, los primeros cinco éxitos constituyen una subsecuencia máxima de 5 elementos semejantes (esto es, incluye el número máximo de elementos semejantes antes de encontrar un F). (Los primeros 4 elementos forman una subsecuencia de elementos semejantes, pero no es máxima porque el quinto elemento también podría estar incluido.) En consecuencia, los 20 elementos están dispuestos en cinco corridas de ensayo, el primero conteniendo cinco S, el segundo conteniendo dos F, y así sucesivamente.

Un número muy grande o muy pequeño de corridas de ensayo en una secuencia indica no aleatoriedad. Por tanto, sea R (el número de corridas de ensayo en una secuencia) el estadístico de prueba y sean $R \leq k_1$ y $R \geq k_2$ la región de rechazo, como se indica en la Figura 15.3. Entonces debemos hallar la distribución de probabilidad para R, $P(R = r)$, para calcular α y localizar una región de rechazo aceptable para la prueba.

Suponga que la secuencia completa contiene n_1 elementos S y n_2 elementos F, resultando en Y_1 corridas de ensayo de las S y Y_2 corridas de ensayo de las F, donde $(Y_1 + Y_2) = R$. Entonces, para una Y_1 determinada, Y_2 puede ser igual a Y_1, $(Y_1 - 1)$ o $(Y_1 + 1)$. Con m denote el número máximo posible de corridas de ensayo. Observe que $m = 2n_1$ si $n_1 = n_2$ y que $m = (2n_1 + 1)$ si $n_1 < n_2$. Supondremos que todo arreglo distinguible de los $(n_1 + n_2)$ elementos de la secuencia constituye un evento simple para el experimento y que los puntos muestrales son igualmente probables. Nos resta a continuación contar el número de puntos muestrales que implican R corridas.

El número total de arreglos distinguibles de n_1 elementos S y n_2 elementos F es

$$\binom{n_1 + n_2}{n_1},$$

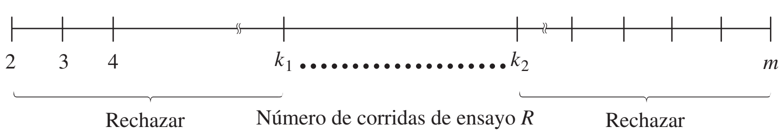


FIGURA **15.3** Región de rechazo para la prueba de corridas de ensayo

FIGURA **15.4** Distribución de n_1 elementos S en y_1 celdas (ninguna vacía)

$$|S|SSSS|SS\ldots|SS|SSS|S|$$

y por tanto la probabilidad por punto muestral es

$$\frac{1}{\binom{n_1+n_2}{n_1}}.$$

El número de formas de lograr y_1 corridas de ensayo S es igual al número de arreglos identificables de n_1 elementos indistinguibles en y_1 celdas, ninguna de las cuales está vacía, como se representa en la Figura 15.4. Esto es igual al número de formas de distribuir $(y_1 - 1)$ barras interiores en los $(n_1 - 1)$ espacios entre los S elementos (las dos barras exteriores permanecen fijas). En consecuencia, es igual al número de formas de seleccionar $(y_1 - 1)$ espacios (para las barras) fuera de los $(n_1 - 1)$ espacios disponibles, o sea

$$\binom{n_1-1}{y_1-1}.$$

El número de formas de observar y_1 corridas de ensayo S y y_2 corridas de ensayo F, obtenido al aplicar la regla *mn*, es

$$\binom{n_1-1}{y_1-1}\binom{n_2-1}{y_2-1}.$$

Esto da el número de puntos muestrales en el evento "y_1 corridas de ensayo de las S y y_2 corridas de ensayo de las F". Entonces, al multiplicar este número por la probabilidad por punto muestral, obtenemos la probabilidad de exactamente y_1 corridas de ensayo de las S y y_2 corridas de ensayo de las F:

$$p(y_1, y_2) = \frac{\binom{n_1-1}{y_1-1}\binom{n_2-1}{y_2-1}}{\binom{n_1+n_2}{n_1}}.$$

Entonces, $P(R = r)$ es igual a la suma de $p(y_1, y_2)$ en todos los valores de y_1 y y_2 tales que $(y_1 + y_2) = r$.

Para ilustrar el uso de la fórmula, el evento $R = 4$ podría ocurrir cuando $y_1 = 2$ y $y_2 = 2$ con elementos S o F comenzando las secuencias. Por tanto,

$$P(R = 4) = 2P(Y_1 = 2, Y_2 = 2).$$

Por otra parte, $R = 5$ podría ocurrir cuando $y_1 = 2$ y $y_2 = 3$ o cuando $y_1 = 3$ y $y_2 = 2$, y estos sucesos son mutuamente excluyentes. Entonces,

$$P(R = 5) = P(Y_1 = 3, Y_2 = 2) + P(Y_1 = 2, Y_2 = 3).$$

EJEMPLO **15.9** Suponga que una secuencia está formada por $n_1 = 5$ elementos S y $n_2 = 3$ elementos F. Calcule la probabilidad de observar $R = 3$ corridas de ensayo. Del mismo modo, calcule $P(R \leq 3)$.

Solución Tres corridas de ensayo podrían ocurrir cuando $y_1 = 2$ y $y_2 = 1$, o cuando $y_1 = 1$ y $y_2 = 2$. Entonces

$$P(R = 3) = P(Y_1 = 2, Y_2 = 1) + P(Y_1 = 1, Y_2 = 2)$$
$$= \frac{\binom{4}{1}\binom{2}{0}}{\binom{8}{5}} + \frac{\binom{4}{0}\binom{2}{1}}{\binom{8}{5}} = \frac{4}{56} + \frac{2}{56} = .107.$$

A continuación requerimos que $P(R \leq 3) = P(R = 2) + P(R = 3)$. De conformidad con esto,

$$P(R = 2) = 2P(Y_1 = 1, Y_2 = 1) = (2)\frac{\binom{4}{0}\binom{2}{0}}{\binom{8}{5}} = \frac{2}{56} = .036.$$

Así, la probabilidad de 3 o menos corridas de ensayo es.107 + .036 = .143. ■

Los valores de $P(R \leq \alpha)$ se dan en la Tabla 10, Apéndice 3, para todas las combinaciones de n_1 y n_2, donde n_1 y n_2 son menores o iguales a 10. Éstas se pueden usar para localizar las regiones de rechazo de pruebas de una o de dos colas. Ilustramos con un ejemplo.

EJEMPLO **15.10** Un examen de verdadero–falso se construyó con las preguntas que corren en la siguiente secuencia:

$$T \quad F \quad F \quad T \quad F \quad T \quad F \quad T \quad T \quad F \quad T \quad F \quad F \quad T \quad F \quad T \quad F \quad T \quad T \quad F.$$

(T = verdadero; F = falso)

¿Esta secuencia indica una desviación de la aleatoriedad en el arreglo de respuestas T y F?

Solución La secuencia contiene $n_1 = 10$ respuestas T y $n_2 = 10$ respuestas F, con $y = 16$ corridas de ensayo. La no aleatoriedad puede estar indicada ya sea por un número inusualmente pequeño o grande de corridas de ensayo; en consecuencia, estaremos usando una prueba de dos colas.

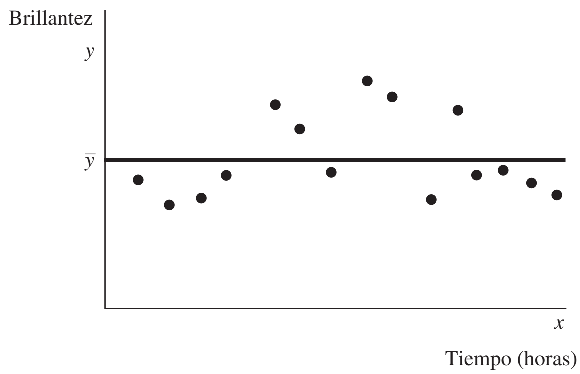
Suponga que deseamos usar α aproximadamente igual a .05 con .025 o menos en cada cola de la región de rechazo. Entonces, de la Tabla 10, Apéndice 3, con $n_1 = n_2 = 10$, vemos que $P(R \leq 6) = .019$ y $P(R \leq 15) = .981$. Entonces, $P(R \geq 16) = 1 - P(R \leq 15) = .019$ y rechazaríamos la hipótesis de aleatoriedad en el nivel de significancia de $\alpha = .038$ si $R \leq 6$ o $R \geq 16$. Como $R = 16$ para los datos observados, concluimos que existe evidencia para indicar no aleatoriedad en el arreglo de las respuestas del profesor. El intento de mezclar las respuestas está exagerado. ■

Una segunda aplicación de la prueba de corridas de ensayo está en detectar no aleatoriedad de una secuencia de medidas cuantitativas en el tiempo. Estas secuencias, conocidas como *series de tiempo*, se presentan en numerosos campos. Por ejemplo, la medición de una característica de calidad de un producto industrial, presión sanguínea de una persona y el precio de una acción en el mercado accionario varían todos en el tiempo. Las desviaciones de la aleatoriedad en una serie, causadas ya sea por tendencias o periodicidades, se pueden detectar al examinar las desviaciones de las mediciones de la serie de tiempo de su promedio. Las desviaciones negativas y positivas podrían estar denotadas por S y F, respectivamente, y podríamos entonces probar esta secuencia de tiempo de desviaciones para no aleatoriedad. Ilustramos con un ejemplo.

EJEMPLO **15.11** Se produce papel en un proceso continuo. Suponga que una medición de la brillantez Y se hace en el papel una vez por hora y que los resultados aparecen como se muestra en la Figura 15.5.

El promedio $\overline{y}$ para las 15 mediciones muestrales aparece en la figura. Observe las desviaciones arriba de $\overline{y}$. ¿Estos datos indican falta de aleatoriedad y por tanto sugieren periodicidad y falta de control en el proceso?

FIGURA **15.5** Brillantez del papel contra tiempo



Solución La secuencia de desviaciones negativas (S) y positivas (F), como se indica en la Figura 15.5, es

$$S \quad S \quad S \quad S \quad F \quad F \quad S \quad F \quad F \quad S \quad F \quad S \quad S \quad S \quad S.$$

Entonces, $n_1 = 10$, $n_2 = 5$ y $R = 7$. Consultando la Tabla 10 del Apéndice 3, encontramos $P(R \leq 7) = .455$. Este valor de R no es improbable, suponiendo que la hipótesis de aleatoriedad sea verdadera. En consecuencia, no hay suficiente evidencia para indicar no aleatoriedad en la secuencia de mediciones de brillantez. ■

La prueba de corridas de ensayo también se puede usar para comparar dos distribuciones de frecuencia poblacional para un experimento de dos muestras no pareado. De esta forma, tenemos una alternativa para la prueba U de Mann–Whitney (Sección 15.6). Si las mediciones para las dos muestras se arreglan en orden de magnitud, forman una secuencia. Las mediciones para las muestras 1 y 2 se pueden denotar como S y F, respectivamente y una vez más nos ocupamos de una prueba de aleatoriedad. Si todas las mediciones para la muestra 1 son más pequeñas que las de la muestra 2, la secuencia resultará en $SSSS \ldots SFFF \ldots F$ o $R = 2$ corridas de ensayo. Un valor pequeño de R proporciona evidencia de una diferencia en las distribuciones poblacionales de frecuencia y la región de rechazo escogida es $R \leq a$. Esta región de rechazo implica una prueba estadística de una cola. Un ejemplo de la aplicación de la prueba de corridas de ensayo para comparar dos distribuciones poblacionales de frecuencia se deja como ejercicio.

Al igual que en el caso de los otros estadísticos de prueba no paramétricos estudiados en secciones anteriores de este capítulo, la distribución de probabilidad para R tiende hacia la normalidad cuando n_1 y n_2 se hacen grandes. La aproximación es buena cuando n_1 y n_2 sean ambos mayores que 10. Por esta razón, podemos usar el estadístico Z como estadístico de prueba con muestras grandes, donde

$$Z = \frac{R - E(R)}{\sqrt{V(R)}},$$

y

$$E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1,$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

son el valor esperado y la varianza de R, respectivamente. La región de rechazo para una prueba de dos colas, con $\alpha = .05$, es $|z| \geq 1.96$. Si α es la probabilidad deseada de un error tipo I, para una prueba de cola superior, rechazamos la hipótesis nula si $z > z_\alpha$ (para una prueba de cola inferior, rechazamos H_0 si $z < -z_\alpha$).

Ejercicios

15.46 Considere una prueba de corridas de ensayo basado en $n_1 = n_2 = 5$ elementos. Suponiendo que H_0 sea verdadera, use la Tabla 10, Apéndice 3, para hallar lo siguiente:

a $P(R = 2)$.

b $P(R \leq 3)$.

c $P(R \leq 4)$.

15.47 El supervisor de un sindicato dice que quienes solicitan empleos son seleccionados sin considerar de qué raza son. Los registros de contratación del local, que contiene sólo miembros masculinos, mostraron la siguiente secuencia de contratación de blancos (W) y negros (B):

$$W \quad W \quad W \quad W \quad B \quad W \quad W \quad W \quad B \quad B \quad W \quad B \quad B$$

¿Estos datos sugieren una selección racial no aleatoria en la contratación de miembros del sindicato?

15.48 Las condiciones (D por enfermo, S por sano) de los árboles individuales de una fila de diez álamos se encontró que estaban, de izquierda a derecha:

$$S \quad S \quad D \quad D \quad S \quad D \quad D \quad D \quad S \quad S$$

¿Hay suficiente evidencia para indicar no aleatoriedad en la secuencia y por tanto la posibilidad de contagio?

15.49 Las piezas que salen de un proceso continuo de producción se clasificaron como defectuosas (D) o no defectuosas (N). Una secuencia de piezas observada en el tiempo fue como sigue:

$$D \quad N \quad N \quad N \quad N \quad N \quad N \quad D \quad D \quad N \quad N \quad N \quad N \quad N \quad N \quad D \quad D$$
$$D \quad N \quad N \quad N \quad N \quad N \quad D \quad N \quad N \quad N \quad D \quad D \quad N \quad N \quad N \quad D \quad D.$$

a Calcule la probabilidad de que $R \leq 11$, donde $n_1 = 11$ y $n_2 = 23$.

b ¿Estos datos sugieren falta de aleatoriedad en la presencia de piezas defectuosas y no defectuosas? Use la aproximación de muestra grande para la prueba de corridas de ensayo.

15.50 Una gráfica de control de calidad se ha mantenido para una característica medible de piezas tomadas de una banda transportadora en un punto fijo de una línea de producción. Las mediciones obtenidas hoy, en orden de tiempo, son como sigue:

68.2	71.6	69.3	71.6	70.4	65.0	63.6	64.7
65.3	64.2	67.6	68.6	66.8	68.9	66.8	70.1

a Clasifique las mediciones de esta serie de tiempo como arriba o abajo de la media muestral y determine (usando la prueba de corridas de ensayo) si observaciones consecutivas sugieren falta de estabilidad en el proceso de producción.

b Divida el periodo en dos partes iguales y compare las medias usando la prueba t de Student. ¿Los datos dan evidencia de un cambio en el nivel medio de las características de calidad? Explique.

15.51 Consulte el Ejercicio 15.24. Use la prueba de corridas de ensayo para analizar los datos. Compare su respuesta a este ejercicio con su respuesta al Ejercicio 15.24.

15.52 Consulte el Ejercicio 15.25. Si en verdad las baterías experimentales tienen una vida media más larga, ¿cuál sería el efecto de esto en el número esperado de corridas de ensayo? Usando la teoría de muestras grandes para la prueba de corridas de ensayo, pruebe (usando $\alpha = .05$) si hay una diferencia en las distribuciones de la duración de baterías para las dos poblaciones. Proporcione el valor p aproximado.

15.10 Coeficiente de correlación de rangos

En las secciones anteriores usamos rangos para indicar la magnitud relativa de observaciones en pruebas no paramétricas para comparación de tratamientos. Ahora empleamos la misma técnica para probar una correlación entre dos variables clasificadas. Dos coeficientes de correlación de rango común son el estadístico r_S de Spearman y el τ de Kendall. Presentamos el r_S de Spearman porque su cálculo es análogo al del coeficiente r de correlación muestral de la Sección 11.8. El coeficiente de correlación de rango de Kendall se estudia con detalle en la obra de Kendall y Stuart (1979).

Suponga que ocho profesores de ciencias elementales han sido clasificados por un juez según su capacidad de enseñanza y todos han hecho un examen nacional de profesores. Los datos se presentan en la Tabla 15.8. ¿Los datos sugieren concordancia entre la clasificación del juez y la calificación del examen? Alternativamente podríamos expresar esta pregunta al consultar si existe correlación entre la clasificación del juez y los rangos de calificaciones del examen.

Las dos variables de interés son la clasificación y la calificación del examen. La primera ya está en forma de rango y las calificaciones del examen se pueden clasificar de modo análogo, como se muestra entre paréntesis en la Tabla 15.8. Las clasificaciones para observaciones em-

Tabla 15.8 Datos para profesores de ciencias

Profesor	Clasificación del juez	Calificación del examen
1	7	44 (1)
2	4	72 (5)
3	2	69 (3)
4	6	70 (4)
5	1	93 (8)
6	3	82 (7)
7	8	67 (2)
8	5	80 (6)

patadas se obtienen al promediar las clasificaciones que ocuparían las observaciones empatadas, como se hace para el estadístico U de Mann–Whitney.

Recuerde que el coeficiente de correlación muestral (Sección 11.8) para observaciones $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ está dado por

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]}}.$$

Con $R(x_i)$ denote el rango de x_i entre $x_1, x_2, \ldots, x_n$ y con $R(y_i)$ denote el rango de y_i entre $y_1, y_2, \ldots y_n$. El *coeficiente de correlación de rango de Spearman,* r_S, se calcula al sustituir los rangos como las mediciones pareadas en la fórmula anterior. Así,

$$r_S = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - \frac{1}{n}\left[\sum_{i=1}^{n} R(x_i)\right]\left[\sum_{i=1}^{n} R(y_i)\right]}{\sqrt{\left\{\sum_{i=1}^{n}[R(x_i)]^2 - \frac{1}{n}\left[\sum_{i=1}^{n} R(x_i)\right]^2\right\}\left\{\sum_{i=1}^{n}[R(y_i)]^2 - \frac{1}{n}\left[\sum_{i=1}^{n} R(y_i)\right]^2\right\}}}.$$

Cuando no hay empates en las observaciones x ni en las observaciones y, esta expresión para r_S se reduce algebraicamente a una expresión más sencilla:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \qquad \text{donde } d_i = R(x_i) - R(y_i).$$

Si el número de empates es pequeño en comparación con el número de pares de datos, resultará un error pequeño por usar esta fórmula de atajo. Dejamos la prueba de esta simplificación como ejercicio (Ejercicio 15.78) e ilustramos el uso de la fórmula mediante un ejemplo.

EJEMPLO **15.12** Calcule r_S para la clasificación del juez y para los datos de calificación del examen de la Tabla 15.8.

Solución Las diferencias y cuadrados de las diferencias entre las dos clasificaciones se muestran en la Tabla 15.9.

Sustituyendo en la fórmula para r_S, obtenemos

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(144)}{8(64-1)} = -.714.$$

■

Tabla 15.9 Datos y cálculos para el Ejemplo 15.12

Profesor	$R(x_i)$	$R(y_i)$	d_i	d_i^2
1	7	1	6	36
2	4	5	−1	1
3	2	3	−1	1
4	6	4	2	4
5	1	8	−7	49
6	3	7	−4	16
7	8	2	6	36
8	5	6	−1	1
Total				144

El coeficiente de correlación de rango de Spearman se puede emplear como estadístico de prueba para probar la hipótesis de que no hay asociación entre dos poblaciones. Suponemos que los n pares de observaciones (x_i, y_i) se han seleccionado al azar y, por tanto, la ausencia de cualquier asociación entre las poblaciones implica una asignación aleatoria de los n rangos dentro de cada muestra. Cada asignación aleatoria (para las dos muestras) representa un punto muestral asociado con el experimento y un valor de r_S se puede calcular para cada uno. Es posible calcular la probabilidad de que r_S tome un valor absoluto grande debido sólo a la casualidad y, por tanto, sugiere una asociación entre poblaciones aun cuando exista ninguna.

La región de rechazo para una prueba de dos colas incluye valores de r_S cercanos a $+1$ y a -1. Si la alternativa es que la correlación entre X y Y es negativa, rechazamos H_0 para valores de r_S cercanos a -1. De manera similar, si la alternativa es que la correlación entre X y Y es positiva, rechazamos H_0 para valores positivos grandes de r_S.

Los valores críticos de r_S se dan en la Tabla 11, Apéndice 3. En el renglón superior de la tabla aparecen valores de α que se podrían usar para una prueba de una cola de la hipótesis nula de no asociación entre X y Y. El número de pares de rango n aparece en el lado izquierdo de la tabla. Las entradas en la tabla muestran el valor crítico r_0 para una prueba de una cola. Entonces, $P(r_S \geq r_0) = \alpha$. Por ejemplo, supongamos que usted tiene $n = 8$ pares de rangos y la hipótesis de investigación es que la correlación entre los rangos es positiva. Entonces, usted desea rechazar la hipótesis nula de no asociación sólo para valores positivos grandes de r_S y usará una prueba de una cola. Consultando la Tabla 11 y usando la fila correspondiente a $n = 8$ y la columna para $\alpha = .05$, se lee $r_0 = .643$. Por tanto, rechaza H_0 para todos los valores de r_S mayores o iguales a .643.

Si se desea dar el valor p asociado con un valor observado de $r = .82$, la Tabla 11 dice que H_0 sería rechazada con $\alpha = .025$ pero no con $\alpha = .01$. Así, $.01 <$ valor $p < .025$.

Si usted pretende probar la hipótesis alternativa que afirma que los rangos están correlacionados en forma negativa, la prueba se lleva a cabo exactamente de la misma manera. La única diferencia es que se rechazará la hipótesis nula sí $r_S \leq -.643$. Es decir, sólo se coloca un signo menos frente al valor tabulado de r_0 para obtener el valor crítico de la cola inferior. De manera similar, si $r = -.82$, entonces $.01 <$ valor $p < .025$.

Para realizar una prueba de dos colas, se rechaza la hipótesis nula si $r_S \geq r_0$ o $r_S \leq -r_0$. El valor de α para la prueba es el doble del que se muestra en la parte superior de la tabla. Por ejemplo, si $n = 8$ y se elige la columna .025, se rechaza H_0 si $r_S \geq .738$ o $r_S \leq -.738$. El valor α para la prueba es $2(.025) = 005$.

El valor p asociado con una prueba de dos colas basada en un valor observado de $r = .82$ es el doble (debido a las dos colas) del valor p de una cola; esto es, $.02 <$ valor $p < .05$.

EJEMPLO **15.13** Pruebe la hipótesis de que no hay asociación entre poblaciones para el Ejemplo 15.12. Proporcione límites para el valor p asociado.

Solución El valor crítico de r_S para una prueba de una cola con $\alpha = .05$ y $n = 8$ es .643. Supongamos que no puede ser posible que una correlación entre la clasificación del juez y los rangos de calificaciones del examen de profesores sea positiva. (Rango bajo significa buena enseñanza y debe estar asociado con una alta calificación del examen si el juez y el examen miden la capacidad de enseñanza.) La hipótesis alternativa es que el coeficiente de correlación de rango poblacional ρ_S es menor que cero, de modo que estamos interesados en una prueba estadística de una cola. Entonces, α para la prueba es el valor tabulado .05 y rechazamos la hipótesis nula si $r_S \leq -.643$.

El valor calculado del estadístico de prueba, $r_S = -.714$, es menor que el valor crítico para $\alpha = .05$. Como H_0 es rechazada para $\alpha = .05$ pero no para $\alpha = .025$, el valor p asociado con la prueba se encuentra en el intervalo $.025 < \text{valor } p < .05$. Por tanto, la hipótesis nula es rechazada con un nivel de significancia $\alpha = .05$. Parece que existe algún acuerdo entre las clasificaciones del juez y las calificaciones del examen, pero este acuerdo podría existir aun cuando *ninguna* proporcione un criterio adecuado para medir la capacidad de enseñanza. Por ejemplo, la asociación podría existir si el juez y quienes formularon el examen de los maestros tuvieran un concepto completamente erróneo, pero semejante, de las características de buena enseñanza. ■

Prueba de correlación de rango de Spearman

Hipótesis nula: H_0: no hay asociación entre los pares de rangos.

Hipótesis alternativa : (1) H_a: hay asociación entre los pares de rangos (una prueba de dos colas),

o bien, (2) la correlación entre los pares de rangos es positiva (o negativa) (una prueba de una cola).

Estadístico de prueba:

$$r_S = \frac{n\sum_{i=1}^{n} R(x_i)R(y_i) - \left[\sum_{i=1}^{n} R(x_i)\right]\left[\sum_{i=1}^{n} R(y_i)\right]}{\sqrt{\left\{n\sum_{i=1}^{n}[R(x_i)]^2 - \left[\sum_{i=1}^{n} R(x_i)\right]^2\right\}\left\{n\sum_{i=1}^{n}[R(y_i)]^2 - \left[\sum_{i=1}^{n} R(y_i)\right]^2\right\}}},$$

donde $R(x_i)$ y $R(y_i)$ denotan el rango de x_i entre $x_1, x_2, \ldots, x_n$ y y_i entre $y_1, y_2, \ldots, y_n$, respectivamente.

Región de rechazo: Para una prueba de dos colas, rechazar H_0 si $r_S \geq r_0$ o $r_S \leq -r_0$, donde r_0 se da en la Tabla 11, Apéndice 3. Duplique la probabilidad tabulada para obtener el valor α para la prueba de dos colas. Para una prueba de una cola, rechace H_0 si $r_S \geq r_0$ (para una prueba de cola superior) o $r_S \leq -r_0$ (para una prueba de cola inferior). El valor α para una prueba de una cola es el que se muestra en la Tabla 11, Apéndice 3.

Ejercicios

15.53 Se realizó un experimento para estudiar la relación entre las clasificaciones hechas por calificadores de hojas de tabaco y el contenido de humedad de las hojas de tabaco correspondientes. Doce hojas fueron clasificadas por los calificadores en una escala de 1 a 10 y se hicieron las correspondientes mediciones de contenido de humedad en las mismas hojas. Los datos se muestran en la siguiente tabla. Calcule r_S. ¿Los datos aportan suficiente evidencia para indicar una asociación entre la clasificación de los calificadores y el contenido de humedad de las hojas? Explique.

Hoja	Clasificacióndе los calificadores	Contenido de humedad
1	9	.22
2	6	.16
3	7	.17
4	7	.14
5	5	.12
6	8	.19
7	2	.10
8	6	.12
9	1	.05
10	10	.20
11	9	.16
12	3	.09

15.54 Es frecuente que los envasadores de alimentos perecederos usen conservadores para retardar el proceso de descompsición. Una preocupación es que demasiado conservador cambiará el sabor del alimento. Se realizó un experimento usando porciones de productos alimenticios y agregando cantidades variables de conservadores. Para cada porción de alimento se registran el sabor y el tiempo para que el alimento empiece a descomponerse. La clasificación de sabor es el promedio de clasificación de tres jueces, cada uno de los cuales clasificó cada porción de alimento en una escala de 1 (malo) a 5 (bueno). En la siguiente tabla se muestran doce mediciones. Use una prueba no paramétrica para determinar si están correlacionadas las clasificaciones de sabor y los tiempos de descomposición. Proporcione el valor p asociado e indique la conclusión apropiada para una prueba con un nivel $\alpha = .05$.

Porción de alimento	Días hasta su descomposición	Clasificación de sabor
1	30	4.3
2	47	3.6
3	26	4.5
4	94	2.8
5	67	3.3
6	83	2.7
7	36	4.2
8	77	3.9
9	43	3.6
10	109	2.2
11	56	3.1
12	70	2.9

15.55 Una gran empresa selecciona graduados para contratarlos usando entrevistas y un examen de aptitudes psicológicas. Las entrevistas realizadas en la casa matriz de la empresa fueron mucho más costosas que el examen, que fue posible realizar en las universidades. En consecuencia, la oficina de personal estuvo interesada en determinar si las calificaciones del examen estaban correlacionadas con clasificaciones de la entrevista y si los exámenes pudieran ser sustituidos por entrevistas. La idea no era eliminar las entrevistas sino reducir su número. Diez candidatos fueron clasificados durante las entrevistas y luego examinados. Las calificaciones pareadas se muestran en la siguiente tabla.

Sujeto	Calificación de la entrevista	Calificación del examen
1	8	74
2	5	81
3	10	66
4	3	83
5	6	66
6	1	94
7	4	96
8	7	70
9	9	61
10	2	86

a Calcule el coeficiente de correlación de rangos de Spearman r_S. El rango 1 se asigna al candidato considerado como el mejor.

b ¿Los datos presentan suficiente evidencia para indicar que la correlación entre las clasificaciones de la entrevista y las calificaciones del examen es menor que cero? Si existe esa evidencia, ¿podemos decir que podrían usarse exámenes para reducir el número de entrevistas?

15.56 Un experto en ciencias políticas deseaba examinar la relación entre la imagen del elector de un candidato político conservador y la distancia en millas entre el domicilio del elector y el del candidato. Cada uno de 12 electores calificó al candidato en una escala de 1 a 20. Los datos resultantes se muestran en la tabla siguiente.

Elector	Calificación	Distancia
1	12	75
2	7	165
3	5	300
4	19	15
5	17	180
6	12	240
7	9	120
8	18	60
9	3	230
10	8	200
11	15	130
12	4	130

a Calcule el coeficiente de correlación de rangos de Spearman, r_S.

b ¿Estos datos aportan suficiente evidencia para indicar una correlación negativa entre calificación y distancia?

15.57 Consulte el Ejercicio 15.12. Calcule el coeficiente de correlación de rangos de Spearman para estos datos y pruebe $H_0 : \rho_S = 0$ con un nivel de significancia de 10%.

15.58 La información que se presenta en la siguiente tabla mide la rigidez a la flexión y la torsión medida por pruebas de ingeniería para 12 raquetas de tenis.

Raqueta	Rigidez a la flexión	Rigidez a la torsión
1	419	227
2	407	231
3	363	200
4	360	211
5	257	182
6	622	304
7	424	384
8	359	194
9	346	158
10	556	225
11	474	305
12	441	235

a Calcule el valor del coeficiente de correlación de rangos r_S entre rigidez a la flexión y rigidez a la torsión.

b Use la prueba basada en el coeficiente de correlación de rangos para determinar si hay una relación positiva importante entre rigidez a la flexión y rigidez a la torsión. Use $\alpha = .05$.

15.59 Consulte el Ejercicio 11.4. Considere los valores en libros y auditados como variables aleatorias y pruebe la existencia de correlación positiva entre los dos mediante el uso del coeficiente de correlación de rango de Spearman. Establezca límites para el valor p asociado con la prueba.

15.60 Consulte el Ejercicio 11.8. Tratando los valores de flujo y estáticos como variables aleatorias, pruebe la presencia de una correlación entre los dos usando el coeficiente de correlación de rangos de Spearman, con $\alpha = .10$.

15.11 Comentarios generales sobre las pruebas estadísticas no paramétricas

Las pruebas estadísticas no paramétricas, presentadas en páginas anteriores representan sólo unos cuantos de los numerosos métodos estadísticos no paramétricos de inferencia que existen. Un conjunto mucho más grande de procedimientos no paramétricos, junto con ejemplos resueltos, se proporciona en los textos citados en las bibliografías [por ejemplo, consulte la obra de Conover (1999), Hollander y Wolfe (1999) y Daniel (2000)]. Muchos de los procedimientos para pruebas de hipótesis no paramétricos se pueden adaptar para proporcionar puntos asociados y estimadores de intervalos para parámetros de localización y diferencias en éstos. También hay procedimientos no paramétricos para manejar algunos de los problemas inferenciales asociados con el modelo lineal.

Hemos indicado que los procedimientos de prueba no paramétricos son particularmente útiles cuando las observaciones experimentales son susceptibles de ordenarse, pero no pueden ser medidas en una escala cuantitativa. Los procedimientos estadísticos paramétricos raras veces se pueden aplicar a este tipo de datos. Por tanto, cualesquier procedimientos inferenciales deben estar basados en métodos no paramétricos.

Una segunda aplicación de los métodos estadísticos no paramétricos tiene que ver con la prueba de hipótesis asociada con poblaciones de datos cuantitativos, cuando exista incertidumbre respecto a la satisfacción de suposiciones acerca de la forma de las distribuciones poblacionales. ¿Exactamente qué tan útiles son los métodos no paramétricos para esta situación? Los métodos estadísticos no paramétricos son rápidos y con frecuencia llevan a una decisión inmediata al probar hipótesis. Cuando las condiciones experimentales se desvíen sustancialmente de las suposiciones básicas que son el fundamento de las pruebas paramétricas, las mediciones de respuesta a veces pueden ser transformadas para aliviar la condición, pero con frecuencia aparece una consecuencia desafortunada: la respuesta transformada ya no es significativa desde un punto de vista práctico y el análisis de los datos transformados ya no responde a los objetivos del experimentador. El uso de métodos no paramétricos a veces evita esta dificultad. Por último, observe que innumerables métodos no paramétricos son casi tan eficientes como sus similares paramétricos cuando las suposiciones que sirven de base a los procedimientos paramétricos son verdaderas; y, como ya antes observamos, podrían ser más eficientes cuando las suposiciones no se satisfacen. Estas razones sugieren que las técnicas no paramétricas desempeñan un papel muy útil en la metodología estadística.

Bibliografía y lecturas adicionales

Conover, W. J. 1999. *Practical Nonparametric Statistics*, 3d ed. New York: Wiley.

Daniel, W. W. 2000. *Applied Nonparametric Statistics*, 2d ed. Pacific Grove, Calif.: Duxbury.

Friedman, M. 1937. “The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance,” *Journal of the American Statistical Association* 32: 675–701.

Gibbons, J. D., and S. Chakraborti. 2003. *Nonparametric Statistical Inference*, 4th ed. New York: Dekker.

Hajek, J., Z. Sidek, and P. K. Sen. 1999. *Theory of Rank Tests*. San Diego: Academic Press.

Hollander, M., and D. A. Wolfe. 1999. *Nonparametric Statistical Methods*, 2d ed. New York: Wiley.

Kendall, M. G., and A. Stuart. 1979. *The Advanced Theory of Statistics*, 4th ed., vol 2. New York: Hafner Press.

Kruskal, W. H., and W. A. Wallis. 1952. “Use of Ranks in One-Criterion Variance Analysis”, *Journal of the American Statistical Association* 47: 583–621.

Lehmann, E. L., and H. J. M. D’Abrera. 2006. *Nonparametrics: Statistical Methods Based on Ranks*. New York: Springer.

Mann, H. B., and Whitney, D. R. 1947. “On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other”, *Annals of Mathematical Statistics* 18: 50–60.

Siegel, S. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Wald, A., and J. Wolfowitz. 1940. “On a Test Whether Two Samples Are from the Same Population”, *Annals of Mathematical Statistics* 2: 147–162.

Wasserman, L. 2006. *All of Nonparametric Statistics*. New York: Springer.
Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods", *Biometrics* 1: 80–83.

Ejercicios complementarios

15.61 A medida que los consumidores se interesan cada vez más en tomar alimentos sanos, muchos productos "de dieta", "sin grasa" o "sin colesterol" están apareciendo en el mercado. A cinco miembros de un jurado, todos ellos expertos en nutrición y elaboración de alimentos, se les pidió que clasificaran tres sustitutos de huevo con bajo contenido de colesterol con base en el gusto, aspecto, textura y si ellos comprarían los productos.[7] Los datos de la siguiente tabla son las calificaciones de sabor (0 = bajo a 20 = alto) de los jueces para los tres productos.

Calificador	Saludable	Mezclados	Huevos revueltos
Dan Bowe	16	9	7
John Carroll	16	7	8
Donna Katzl	14	8	4
Rick O'Connell	15	16	9
Roland Passot	13	11	2

a ¿Qué diseño se utilizó en este experimento para probar el sabor?

b ¿Piensa usted que es probable que los datos satisfagan las suposiciones necesarias para poner en práctica una teoría normal basada en ANOVA?

c ¿Los datos indican que hay diferencias importantes entre las respuestas para los diferentes sustitutos de huevo? Determine o proporcione un límite al valor p aproximado.

15.62 Dos gastrónomos, A y B, calificaron 20 comidas en una escala de 1 a 10. Los datos se muestran en la siguiente tabla. ¿Los datos dan suficiente evidencia para indicar que uno de los gastrónomos tiende a dar calificaciones más altas que el otro? Pruebe usando la prueba de signos con un valor de α cercano a .05.

Comida	A	B	Comida	A	B
1	6	8	11	6	9
2	4	5	12	8	5
3	7	4	13	4	2
4	8	7	14	3	3
5	2	3	15	6	8
6	7	4	16	9	10
7	9	9	17	9	8
8	7	8	18	4	6
9	2	5	19	4	3
10	4	3	20	5	5

15.63 Refiérase a la comparación de calificaciones aplicadas por gastrónomos a las comidas del Ejercicio 15.62 y use la prueba de rangos con signos de Wilcoxon para determinar si los datos dan suficiente evidencia para indicar una diferencia en las calificaciones de los dos gastrónomos. Pruebe usando un valor de α cercano a .05. Compare los resultados de esta prueba con los de la prueba de signos del Ejercicio 15.62. ¿Las conclusiones son consistentes?

7. *Fuente*: Karola Sakelel, "Egg Substitutes Range in Quality," *San Francisco Chronicle,* 10 de febrero de 1993, p. 8.

15.64 En una investigación de la conducta de exploración visual de niños sordos, se tomaron mediciones de la rapidez de movimiento de los ojos de nueve niños sordos y de nueve que sí escuchan. A partir de los datos dados en la tabla, ¿hay suficiente evidencia para justificar la afirmación de que las distribuciones de la rapidez de movimiento de los ojos difieren para niños sordos A y niños B que sí escuchan?

	Niños sordos A	Niños que sí escuchan B
	2.75 (15)	.89 (1)
	2.14 (11)	1.43 (7)
	3.23 (18)	1.06 (4)
	2.07 (10)	1.01 (3)
	2.49 (14)	.94 (2)
	2.18 (12)	1.79 (8)
	3.16 (17)	1.12 (5.5)
	2.93 (16)	2.01 (9)
	2.20 (13)	1.12 (5.5)
Suma de rangos	126	45

15.65 Una comparación de reacción (en segundos) a dos estímulos diferentes en un experimento psicológico de asociación de palabras produjo los resultados de la siguiente tabla cuando se aplicó a una muestra aleatoria de 16 personas. ¿Los datos presentan suficiente evidencia para indicar una diferencia en la localización para las distribuciones de tiempos de reacción para los dos estímulos? Use el estadístico U de Mann–Whitney y pruebe con $\alpha = .05$. (*Nota*: esta prueba fue realizada usando la prueba t de Student en el Ejercicio 13.3. Compare sus resultados.)

Estímulo 1	Estímulo 2
1	4
3	2
2	3
1	3
2	1
1	2
3	3
2	3

15.66 Si (como en el caso de mediciones producidas por dos instrumentos bien calibrados) las medias de dos poblaciones son iguales, el estadístico U de Mann–Whitney se puede usar para probar hipótesis relacionadas con varianzas poblacionales (o medidas de variabilidad más generales) de la siguiente manera. Al igual que en la Sección 15.6, identifique la población I como aquella de la cual se toma la muestra más pequeña. Clasifique la muestra combinada. Numere las observaciones clasificadas de afuera hacia adentro; esto es, el número de observación más pequeño, 1; el más grande, 2; el siguiente al más pequeño, 3; el siguiente al más grande, 4; y así sucesivamente. Esta sucesión final de números induce un ordenamiento en los símbolos x (observaciones de la muestra I) y y (observaciones de la muestra II). Si $\sigma_X^2 < \sigma_Y^2$, uno esperaría hallar una predominancia de las x con rangos altos y por tanto una suma relativamente grande de rangos para las observaciones x. Por el contrario, si $\sigma_X^2 > \sigma_Y^2$, casi todas las x tendrían rangos bajos y la suma de los rangos de las observaciones x sería pequeña.

a Dadas las mediciones de la siguiente tabla, producidas por instrumentos de precisión bien calibrados, A y B, pruebe cerca del nivel $\alpha = .05$ para determinar si el instrumento más costoso B es más preciso que el instrumento A. (Observe que esto implica una prueba de una cola.) Use la prueba U de Mann–Whitney.

Instrumento A	Instrumento B
1060.21	1060.24
1060.34	1060.28
1060.27	1060.32
1060.36	1060.30
1060.40	

b Pruebe usando el estadístico F de la Sección 10.9.

15.67 Calcule la probabilidad de que $U \leq 2$ para $n_1 = n_2 = 5$. Suponga que no hay empates y que H_0 es verdadera.

15.68 Calcule la probabilidad de que la prueba T de Wilcoxon (Sección 15.4) es menor o igual a 2 para $n = 3$ pares. Suponga que no hay empates y que H_0 es verdadera.

15.69 Para investigar posibles diferencias entre ritmos de producción para tres líneas de producción que producen piezas similares, unos examinadores tomaron muestras aleatorias independientes de las cifras totales de producción durante 7 días para cada línea. Los datos resultantes aparecen en la siguiente tabla. ¿Estos datos aportan suficiente evidencia para indicar algunas diferencias de localización para los tres conjuntos de cifras de producción, con un nivel de significancia de 5%?

Línea 1	Línea 2	Línea 3
48	41	18
43	36	42
39	29	28
57	40	38
21	35	15
47	45	33
58	32	31

15.70 **a** Suponga que una compañía desea estudiar la forma en que la personalidad se relaciona con el liderazgo. Se seleccionan cuatro supervisores, I, II, III y IV, con diferentes tipos de personalidad. A continuación se seleccionan varios empleados del grupo supervisado por cada uno y a estos empleados se les pide clasificar al líder de su grupo en una escala de 1 a 20 (20 significa altamente favorable). La siguiente tabla muestra los datos resultantes. ¿Hay suficiente evidencia para indicar que uno o más de los supervisores tiende a recibir calificaciones más altas que los otros? Use $\alpha = 0.05$.

I	II	III	IV
20	17	16	8
19	11	15	12
20	13	13	10
18	15	18	14
17	14	11	9
	16		10

b Suponga que la compañía está particularmente interesada en comparar las calificaciones de los tipos de personalidad representados por los supervisores I y III. Haga esta comparación usando $\alpha = .05$.

15.71 Los líderes de un sindicato obrero desean determinar las preferencias de sus miembros antes de negociar con la administración. Diez miembros del sindicato se seleccionan aleatoriamente y cada uno de ellos llenó un extenso cuestionario. Las respuestas a los diversos aspectos del cuestionario harán posible que el sindicato califique, en orden de importancia, los puntos a negociar. Las calificaciones muestrales se ven en la siguiente tabla. ¿Hay suficiente evidencia para indicar que uno o más de los puntos se prefieren a otros? Pruebe usando $\alpha = .05$.

Persona	Más salario	Estabilidad laboral	Prestaciones	Menos horas de trabajo
1	2	1	3	4
2	1	2	3	4
3	4	3	2	1
4	1	4	2	3
5	1	2	3	4
6	1	3	4	2
7	2.5	1	2.5	4
8	3	1	4	2
9	1.5	1.5	3	4
10	2	3	1	4

15.72 Se formaron seis grupos de tres niños con IQ y edad comparables. A cada niño se le enseñó el concepto de tiempo usando uno de tres métodos: lectura, demostración o máquina de enseñanza. Las calificaciones mostradas en la siguiente tabla indican el aprovechamiento de los estudiantes cuando fueron examinados para ver lo bien que habían entendido el concepto. ¿Hay suficiente evidencia para indicar que los métodos de enseñanza difieren en efectividad? Establezca límites para el valor p.

Grupo	Lectura	Demostración	Máquina de enseñanza
1	20	22	24
2	25	25	27
3	30	40	39
4	37	26	41
5	24	20	21
6	16	18	25

15.73 Calcule $P(R \leq 6)$ para la prueba de corridas de ensayo, donde $n_1 = n_2 = 8$ y H_0 es verdadera. No use la Tabla 10, Apéndice 3.

15.74 Considere una prueba de suma de rangos de Wilcoxon para comparar dos distribuciones de probabilidad con base en muestras aleatorias independientes de $n_1 = n_2 = 5$. Encuentre $P(W \leq 17)$, suponiendo que H_0 es verdadera.

***15.75** Para la muestra de la población I, denote con U el estadístico de Mann–Whitney y denote con W el estadístico de suma de rangos de Wilcoxon.[8] Demuestre que

$$U = n_1 n_2 + (1/2)n_1(n_1 + 1) - W.$$

***15.76** Consulte el Ejercicio 15.75.

a Demuestre que $E(U) = (1/2)n_1 n_2$ cuando H_0 es verdadera.

b Demuestre que $V(U) = (1/12)[n_1 n_2(n_1 + n_2 + 1)]$ cuando H_0 es verdadera, donde H_0 expresa que las dos poblaciones tienen distribuciones idénticas.

***15.77** Denote con T el estadístico de rangos con signo de Wilcoxon para n pares de observaciones. Demuestre que $E(T) = (1/4)n(n + 1)$ y $V(T) = (1/24)[n(n + 1)(2n + 1)]$ cuando las dos poblaciones son idénticas. Observe que estas propiedades no dependen de si T se construye a partir de diferencias negativas o positivas.

8. Los ejercicios precedidos por un asterisco son opcionales.

***15.78** Considere el coeficiente de correlación de rangos de Spearman de la Sección 15.10. Demuestre que, cuando no hay empates en cualquiera de las observaciones x o las observaciones y, entonces

$$
\begin{aligned}
r_S &= \frac{n\sum_{i=1}^{n} R(x_i)R(y_i) - \left[\sum_{i=1}^{n} R(x_i)\right]\left[\sum_{i=1}^{n} R(y_i)\right]}{\sqrt{\left\{n\sum_{i=1}^{n}[R(x_i)]^2 - \left[\sum_{i=1}^{n} R(x_i)\right]^2\right\}\left\{n\sum_{i=1}^{n}[R(y_i)]^2 - \left[\sum_{i=1}^{n} R(y_i)\right]^2\right\}}} \\
&= 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},
\end{aligned}
$$

donde $d_i = R(x_i) - R(y_i)$.

CAPÍTULO 16

Introducción a los métodos de Bayes para inferencia

16.1 Introducción

Iniciamos este capítulo con un ejemplo que ilustra los conceptos y una aplicación de la propuesta de Bayes para hacer inferencias. Suponga que estamos interesados en estimar la proporción de quienes responden a una nueva terapia para tratar una enfermedad que es grave y difícil de curar (por ejemplo una enfermedad que se dice es virulenta). Si p denota la probabilidad de que cualquier persona con la enfermedad responda al tratamiento, el número Y de quienes respondan en una muestra de tamaño n podría suponerse razonablemente que tiene una distribución binomial con parámetro p. En capítulos previos hemos visto que el parámetro p tiene un valor fijo pero desconocido y hemos examinado estimadores puntuales, estimadores de intervalo y pruebas de hipótesis para este parámetro. Incluso antes de que recolectemos dato alguno, nuestro conocimiento de que la enfermedad es virulenta podría llevarnos a pensar que es probable que el valor de p sea relativamente pequeño, quizá en la proximidad de .25. ¿Cómo podemos usar esta información en el proceso de hacer inferencias acerca de p?

Una forma de usar esta información previa acerca de p es utilizar un método de Bayes. En esta proposición modelamos la distribución *condicional* de Y dada p, $Y \mid p$, como binomial:

$$p(y \mid p) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n.$$

La incertidumbre acerca del parámetro p se maneja tratándolo como una variable aleatoria y, antes de observar cualquier dato, asignar una distribución *previa* a p. Dado que sabemos que $0 < p < 1$ y que la función de densidad beta tiene un intervalo (0, 1) como apoyo, es conveniente usar una distribución beta como previa para p. Pero, ¿cuál distribución beta deberíamos usar?

Como la media de una variable aleatoria con distribución beta y parámetros α y β es $\mu = \alpha/(\alpha + \beta)$ y pensamos que p podría estar en la proximidad de .25, podríamos escoger usar una distribución beta con $\alpha = 1$ y $\beta = 3$ (y $\mu = .25$) como la previa para p. Por tanto, la densidad asignada a p es

$$g(p) = \frac{1}{3}(1 - p)^2, \qquad 0 < p < 1.$$

Como hemos especificado la distribución condicional de $Y|p$ y la distribución de p, también hemos especificado la distribución conjunta de (Y, p) y podemos determinar la distribución marginal de Y y la distribución condicional de $p|Y$. *Después* de observar $Y = y$, la densidad *posterior* de p dada $Y = y$, $g^*(p|y)$ se puede determinar. En la siguiente sección deducimos un resultado general que, en nuestro ejemplo de una enfermedad virulenta, implica que la densidad posterior de p dada $Y = y$ es

$$g^*(p \mid y) = \frac{\Gamma(n + 4)}{\Gamma(y + 1)\Gamma(n - y + 3)} p^y(1 - p)^{n-y+2}, \qquad 0 < p < 1.$$

Observe que la densidad posterior para $p|y$ es una densidad beta con $\alpha = y + 1$ y $\beta = n - y + 3$. Esta densidad posterior es la densidad "actualizada" (por los datos) de p y es la base para todas las inferencias bayesianas con respecto a p. En las siguientes secciones, describimos el método general de Bayes y especificamos cómo usar la densidad posterior para obtener estimaciones, intervalos creíbles y pruebas de hipótesis para p y para parámetros asociados con otras distribuciones.

16.2 Bayesianos previos, posteriores y estimadores

Si $Y_1, Y_2, \ldots, Y_n$ denotan las variables aleatorias asociadas con una muestra de tamaño n, ya previamente usamos la notación $L(y_1, y_2, \ldots, y_n|\theta)$ para denotar la verosimilitud de la muestra. En el caso discreto, esta función está definida como la probabilidad conjunta $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ y, en el caso continuo, es la densidad conjunta de $Y_1, Y_2, \ldots, Y_n$ evaluada en $y_1, y_2, \ldots, y_n$. El parámetro θ está incluido entre los argumentos de $L(y_1, y_2, \ldots, y_n|\theta)$ para denotar que esta función depende explícitamente del valor de algún parámetro θ. En el método bayesiano, el parámetro desconocido θ se ve como una variable aleatoria con una distribución de probabilidad, llamada *distribución previa* de θ. Esta distribución previa se especifica antes de recolectar cualquier información y da una descripción teórica de la información acerca de θ de la que se disponía antes de obtener cualquier dato. En nuestro análisis inicial supondremos que el parámetro θ tiene una distribución continua con densidad $g(\theta)$ que no tiene parámetros desconocidos.

Usando la probabilidad de los datos y la previa sobre θ, se deduce que la probabilidad conjunta de $Y_1, Y_2, \ldots, Y_n, \theta$ es

$$f(y_1, y_2, \ldots, y_n, \theta) = L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)$$

y que la densidad marginal o función de masa de $Y_1, Y_2, \ldots, Y_n$ es

$$m(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta.$$

Finalmente, la densidad posterior de $\theta|y_1, y_2, \ldots, y_n$ es

$$g^*(\theta \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}.$$

La densidad posterior resume toda la información pertinente acerca del parámetro θ al hacer uso de la información contenida en la densidad previa para θ y la información de los datos.

EJEMPLO **16.1** Denotemos con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Bernoulli donde $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$ y supongamos que la distribución previa para p es beta (α, β). Encuentre la distribución posterior para p.

Solución Como la función de probabilidad de Bernoulli se puede escribir como

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1,$$

la probabilidad $L(y_1, y_2, \ldots, y_n|p)$ es

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid p) &= p(y_1, y_2, \ldots, y_n \mid p) \\ &= p^{y_1}(1-p)^{1-y_1} \times p^{y_2}(1-p)^{1-y_2} \times \cdots \times p^{y_n}(1-p)^{1-y_n} \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i}, \qquad y_i = 0, 1 \;\text{ y }\; 0 < p < 1. \end{aligned}$$

Entonces,

$$\begin{aligned} f(y_1, y_2, \ldots, y_n, p) &= L(y_1, y_2, \ldots, y_n \mid p) \times g(p) \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1} \end{aligned}$$

y

$$\begin{aligned} m(y_1, y_2, \ldots, y_n) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}\, dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(n-\sum y_i+\beta\right)}{\Gamma(n+\alpha+\beta)}. \end{aligned}$$

Finalmente, la densidad posterior de p es

$$\begin{aligned} g^*(p \mid y_1, y_2, \ldots, y_n) &= \frac{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}}{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \dfrac{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(n-\sum y_i+\beta\right)}{\Gamma(n+\alpha+\beta)}}, \qquad 0 < p < 1 \\ &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(n-\sum y_i+\beta\right)} \times \\ &\qquad p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}, \qquad 0 < p < 1, \end{aligned}$$

una densidad beta con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = n - \sum y_i + \beta$. ■

Antes de continuar, veamos algunas de las implicaciones del resultado del Ejemplo 16.1. En el siguiente ejemplo compararemos las distribuciones previa y posterior para algunas (por ahora) elecciones arbitrarias de los parámetros de la distribución previa y los resultados del experimento.

EJEMPLO **16.2** Considere la situación de la enfermedad virulenta y los resultados del Ejemplo 16.1. Compare las distribuciones previa y posterior del parámetro p de Bernoulli (la proporción de quienes responden a la nueva terapia) si escogemos los valores para α y β y observamos la información hipotética dada a continuación:

a $\alpha = 1, \beta = 3, n = 5, \sum y_i = 2$.
b $\alpha = 1, \beta = 3, n = 25, \sum y_i = 10$.
c $\alpha = 10, \beta = 30, n = 5, \sum y_i = 2$.
d $\alpha = 10, \beta = 30, n = 25, \sum y_i = 10$.

Solución Antes de continuar, observe que ambas distribuciones previas beta tienen media

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{1}{1+3} = \frac{10}{10+30} = .25$$

y que ambas muestras hipotéticas resultan en el mismo valor de las estimaciones de probabilidad máxima (MLE) para p:

$$\hat{p} = \frac{1}{n}\sum y_i = \frac{2}{5} = \frac{10}{25} = .40.$$

Como se dedujo en el Ejemplo 16.1, si $y_1, y_2, \ldots, y_n$ denotan los valores en una muestra aleatoria de una distribución de Bernoulli, donde $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$ y la distribución previa para p es beta (α, β), la distribución posterior para p es beta $(\alpha^* = \sum y_i + \alpha,\ \beta^* = n - \sum y_i + \beta)$. Por tanto, para las opciones de este ejemplo,

a cuando la distribución previa es beta (1, 3), $n = 5$, $\sum y_i = 2$, la distribución posterior es beta con

$$\alpha^* = \sum y_i + \alpha = 2 + 1 = 3 \quad \text{y} \quad \beta^* = n - \sum y_i + \beta = 5 - 2 + 3 = 6.$$

b cuando la distribución previa es beta (1, 3), $n = 25$, $\sum y_i = 10$, la distribución posterior es beta con

$$\alpha^* = 10 + 1 = 11 \quad \text{y} \quad \beta^* = 25 - 10 + 3 = 18.$$

c cuando la distribución previa es beta (10, 30), $n = 5$, $\sum y_i = 2$, la distribución posterior es beta con

$$\alpha^* = 2 + 10 = 12 \quad \text{y} \quad \beta^* = 5 - 2 + 30 = 33.$$

d cuando la distribución previa es beta (10, 30), $n = 25$, $\sum y_i = 10$, la distribución posterior es beta con

$$\alpha^* = 20 \quad \text{y} \quad \beta^* = 45.$$

Recuerde que la media y varianza de una variable aleatoria con distribución beta (α, β) son

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Los parámetros de las distribuciones previas y posteriores beta, junto con sus medias y varianzas se resumen en la Tabla 16.1. La figura 16.1(a) contiene gráficas de las distribuciones beta

Tabla 16.1 Distribuciones previas y posteriores beta para el Ejemplo 16.2

Distribución	n	$\sum y_i$	Parámetros de distribución beta	Media	Varianza
Previa	—	—	$\alpha = 1, \quad \beta = 3$	.2500	.0375
Posterior	5	2	$\alpha^* = 3, \quad \beta^* = 6$	.3333	.0222
Posterior	25	10	$\alpha^* = 11, \beta^* = 18$	.4074	.0078
Previa	—	—	$\alpha = 10, \quad \beta = 30$	.2500	.0046
Posterior	5	2	$\alpha^* = 12, \beta^* = 33$	.2667	.0043
Posterior	25	10	$\alpha^* = 20, \beta^* = 45$	.3077	.0032

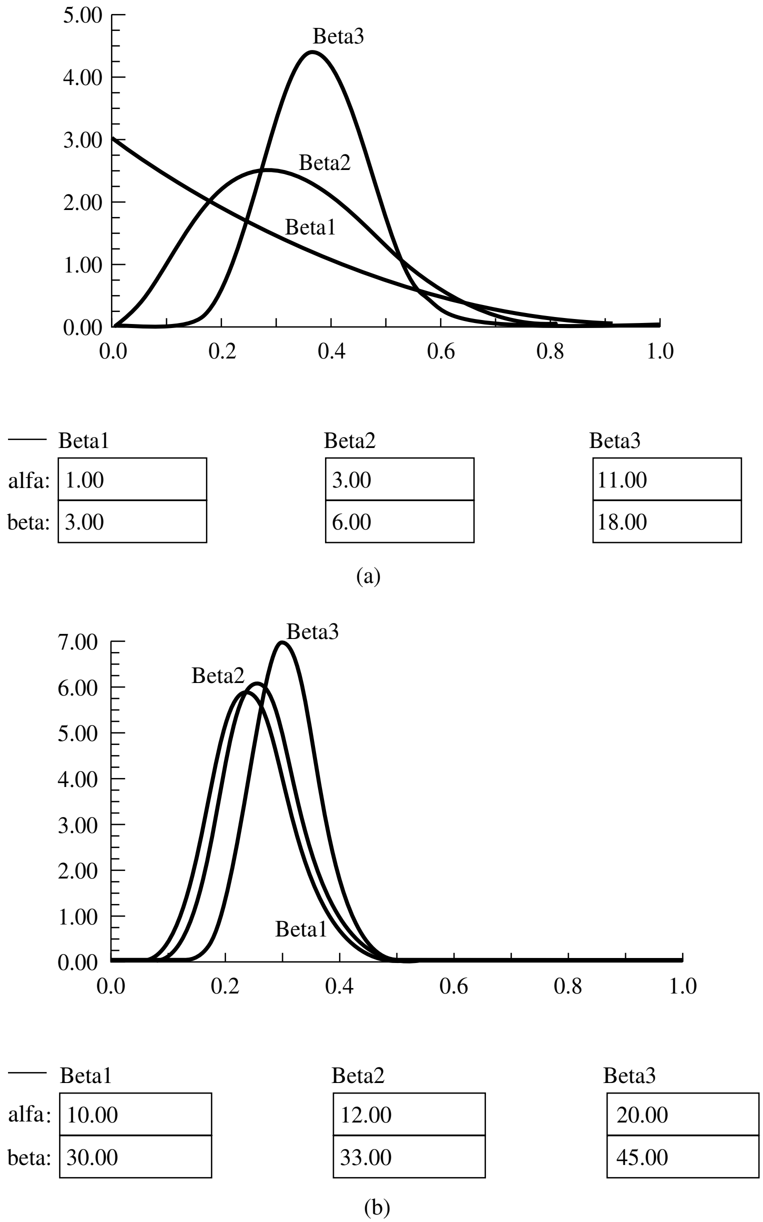


FIGURA **16.1** Gráficas de distribuciones previas y posteriores beta en el (previas y posteriores) asociadas con la previa beta con parámetros $\alpha = 1$, $\beta = 3$. Las gráficas de las distribuciones beta asociadas con la previa beta (10, 30) se dan en la Figura 16.1(b). ■

En los Ejemplos 16.1 y 16.2 obtuvimos densidades posteriores que, como la previa, son densidades beta con valores de parámetro alterados (por los datos).

DEFINICIÓN **16.1**

Las distribuciones previas que resulten en distribuciones posteriores que sean de la misma forma funcional que la previa, pero con valores de parámetro alterados, se denominan *previas conjugadas*.

Cualquier distribución beta es una distribución previa conjugada para una distribución de Bernoulli (o una binomial). Cuando la previa se actualiza (usando los datos), el resultado es una beta posterior con valores de parámetro alterados. Esto es conveniente desde el punto de vista computacional porque podemos determinar la fórmula exacta para la posterior y de ahí en adelante usar propiedades previamente desarrolladas de una distribución conocida. Para las distribuciones que usamos en este capítulo, hay previas conjugadas asociadas con los parámetros relevantes. Es frecuente que estas familias de previas conjugadas se vean lo suficientemente claras para manejar casi todas las situaciones prácticas. Como resultado, las previas conjugadas se usan frecuentemente en la práctica.

Como la posterior es una función de densidad de probabilidad de buena fe, algún resumen característico de esta densidad da una estimación para θ. Por ejemplo, podríamos usar la media, la mediana o la moda de la densidad posterior de nuestro estimador. Si estamos interesados en estimar alguna función de θ, por ejemplo $t(\theta)$, usaremos el valor esperado posterior de $t(\theta)$ como nuestro estimador para esta función de θ.

DEFINICIÓN **16.2**

Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria con función de verosimilitud $L(y_1, y_2, \ldots, y_n | \theta)$, y que θ tenga densidad previa $g(\theta)$. El estimador posterior de Bayes para $t(\theta)$ está dado por

$$\widehat{t(\theta)}_B = E(t(\theta) \mid Y_1, Y_2, \ldots, Y_n).$$

EJEMPLO **16.3** En el Ejemplo 16.1 encontramos la distribución posterior del parámetro p de Bernoulli basado en una previa beta con parámetros (α, β). Encuentre los estimadores de Bayes para p y $p(1 - p)$. [Recuerde que $p(1 - p)$ es la varianza de una variable aleatoria de Bernoulli con parámetro p].

Solución En el Ejemplo 16.1 encontramos que la densidad posterior de p es una densidad beta con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = n - \sum y_i + \beta$:

$$g^*(p \mid y_1, y_2, \ldots, y_n) = \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1 - p)^{\beta^*-1}, \qquad 0 < p < 1.$$

La estimación de p es la media posterior de p. De nuestro estudio previo de la distribución beta, sabemos que

$$\begin{aligned}
\hat{p}_B &= E(p \mid y_1, y_2, \ldots, y_n) \\
&= \frac{\alpha^*}{\alpha^* + \beta^*} \\
&= \frac{\sum y_i + \alpha}{\sum y_i + \alpha + n - \sum y_i + \beta} = \frac{\sum y_i + \alpha}{n + \alpha + \beta}.
\end{aligned}$$

Del mismo modo,

$$\begin{aligned}
[\widehat{p(1-p)}]_B &= E(p(1-p) \mid y_1, y_2, \ldots, y_n) \\
&= \int_0^1 p(1-p)\frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1}\,dp \\
&= \int_0^1 \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*}(1-p)^{\beta^*}\,dp \\
&= \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \times \frac{\Gamma(\alpha^* + 1)\Gamma(\beta^* + 1)}{\Gamma(\alpha^* + \beta^* + 2)} \\
&= \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \times \frac{\alpha^*\Gamma(\alpha^*)\beta^*\Gamma(\beta^*)}{(\alpha^* + \beta^* + 1)(\alpha^* + \beta^*)\Gamma(\alpha^* + \beta^*)} \\
&= \frac{\alpha^*\beta^*}{(\alpha^* + \beta^* + 1)\,(\alpha^* + \beta^*)} \\
&= \frac{\left(\sum y_i + \alpha\right)\left(n - \sum y_i + \beta\right)}{(n + \alpha + \beta + 1)\,(n + \alpha + \beta)}.
\end{aligned}$$

Por tanto, los estimadores de Bayes para p y $p(1-p)$ son

$$\hat{p}_B = \frac{\sum Y_i + \alpha}{n + \alpha + \beta} \quad \text{y} \quad [\widehat{p(1-p)}]_B = \frac{\left(\sum Y_i + \alpha\right)\left(n - \sum Y_i + \beta\right)}{(n + \alpha + \beta + 1)\,(n + \alpha + \beta)}$$ ■

Un examen más a fondo del estimador de Bayes para p dado en el Ejemplo 16.3 da como resultado

$$\begin{aligned}
\hat{p}_B &= \frac{\sum Y_i + \alpha}{n + \alpha + \beta} \\
&= \left(\frac{n}{n + \alpha + \beta}\right)\left(\frac{\sum Y_i}{n}\right) + \left(\frac{\alpha + \beta}{n + \alpha + \beta}\right)\left(\frac{\alpha}{\alpha + \beta}\right) \\
&= \left(\frac{n}{n + \alpha + \beta}\right)\overline{Y} + \left(\frac{\alpha + \beta}{n + \alpha + \beta}\right)\left(\frac{\alpha}{\alpha + \beta}\right).
\end{aligned}$$

Entonces, vemos que el estimador de Bayes para p es un promedio ponderado de la media muestral, $\overline{Y}$ (la MLE para p) y la media de la previa beta asignada a p. Observe que la media previa de p recibe menos valor para tamaños muestrales más grandes, mientras que el valor dado a la media muestral aumenta para tamaños muestrales más grandes. También, como

$E(\overline{Y}) = p$, es fácil ver que el estimador de Bayes para p *no* es un estimador insesgado. En términos generales, los estimadores de Bayes no son insesgados.

Observe que los estimadores obtenidos en el Ejemplo 16.3 son funciones del estadístico suficiente $\sum Y_i$. Ésta no es coincidencia puesto que un estimador de Bayes es siempre una función de un estadístico suficiente, resultado consecuencia del criterio de factorización (vea Teorema 9.4).

Si U es un estadístico suficiente para el parámetro θ basado en una muestra aleatoria Y_1, $Y_2, \ldots, Y_n$, entonces

$$L(y_1, y_2, \ldots, y_n \mid \theta) = k(u, \theta) \times h(y_1, y_2, \ldots, y_n),$$

donde $k(u, \theta)$ es una función sólo de u y θ y $h(y_1, y_2, \ldots, y_n)$ no es una función de θ. Además (vea Hogg, McKean y Craig, 2005), la función $k(u, \theta)$ puede (pero no necesita) ser escogida para ser la función de probabilidad de masa o densidad del estadístico U. De acuerdo con la notación de este capítulo, escribimos la densidad condicional de $U \mid \theta$ como $k(u \mid \theta)$. Entonces, como $h(y_1, y_2, \ldots, y_n)$ no es una función de θ,

$$\begin{aligned}
g^{\star}(\theta \mid y_1, y_2, \ldots, y_n) &= \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta} \\
&= \frac{k(u \mid \theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)}{\int_{-\infty}^{\infty} k(u \mid \theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)\, d\theta} \\
&= \frac{k(u \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} k(u \mid \theta) \times g(\theta)\, d\theta}.
\end{aligned}$$

Por tanto, en casos donde se conozca la distribución de un estadístico U suficiente, la posterior se puede determinar usando la densidad condicional de $U \mid \theta$. Ilustramos con el siguiente ejemplo.

EJEMPLO **16.4** Denotemos con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población normal con media μ desconocida y varianza σ_o^2 *conocida.* La distribución previa conjugada para μ es una distribución normal con media η *conocida* y varianza δ^2 *conocida.* Encuentre la distribución posterior y el estimador de Bayes para μ.

Solución Como $U = \sum Y_i$ es un estadístico suficiente para μ y se sabe que tiene una distribución normal con media $n\mu$ y varianza $n\sigma_o^2$,

$$L(u \mid \mu) = \frac{1}{\sqrt{2\pi n\sigma_o^2}} \exp\left[\frac{1}{2n\sigma_o^2}(u - n\mu)^2\right], \qquad -\infty < u < \infty$$

y la densidad conjunta de U y μ es

$$\begin{aligned}
f(u, \mu) &= L(u \mid \mu) \times g(\mu) \\
&= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}} \exp\left[-\frac{1}{2n\sigma_o^2}(u - n\mu)^2 - \frac{1}{2\delta^2}(\mu - \eta)^2\right], \\
&\qquad -\infty < u < \infty, -\infty < \mu < \infty.
\end{aligned}$$

Veamos la cantidad en el exponente de arriba:

$$
\begin{aligned}
&-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2(u-n\mu)^2+n\sigma_o^2(\mu-\eta)^2\right]\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2u^2-2\delta^2un\mu+\delta^2n^2\mu^2+n\sigma_o^2\mu^2-2n\sigma_o^2\mu\eta+n\sigma_o^2\eta^2\right]\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[(n^2\delta^2+n\sigma_o^2)\mu^2-2(n\delta^2u+n\sigma_o^2\eta)\mu+\delta^2u^2+n\sigma_o^2\eta^2\right]\\
&\quad=-\frac{1}{2\sigma_o^2\delta^2}\left[(n\delta^2+\sigma_o^2)\mu^2-2(\delta^2u+\sigma_o^2\eta)\mu\right]-\frac{1}{2n\sigma_o^2\delta^2}(\delta^2u^2+n\sigma_o^2\eta^2)\\
&\quad=-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left[\mu^2-2\left(\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)\mu+\left(\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]\\
&\qquad\qquad-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2u^2+n\sigma_o^2\eta^2-\frac{n(\delta^2u+\sigma_o^2\eta)^2}{n\delta^2+\sigma_o^2}\right].
\end{aligned}
$$

Finalmente, obtenemos:

$$
\begin{aligned}
-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2&=-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\\
&\qquad-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2.
\end{aligned}
$$

Por tanto,

$$
\begin{aligned}
f(u,\mu)&=\frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2\right]\\
&=\frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]\\
&\qquad\times\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]
\end{aligned}
$$

y

$$
\begin{aligned}
m(u)&=\frac{\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]d\mu\\
&=\frac{\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]}{\sqrt{2\pi n(n\delta^2+\sigma_o^2)}}\int_{-\infty}^{\infty}\frac{\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}\,d\mu.
\end{aligned}
$$

Si reconocemos esta integral como la de una función de densidad normal y por tanto igual a 1, obtenemos que la función de densidad marginal para U es normal con media $n\eta$ y varianza $(n^2\delta^2 + n\sigma_o^2)$. Además, la densidad posterior de μ dada $U = u$ es

$$g^*(\mu \mid u) = \frac{f(u,\mu)}{m(u)} = \frac{1}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}\exp\left[-\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{\delta^2 u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right)^2\right],$$
$$-\infty < \mu < \infty,$$

una densidad normal con media

$$\eta^* = \left(\frac{\delta^2 u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right) \quad \text{y} \quad \text{varianza} \quad \delta^{*2} = \left(\frac{\sigma_o^2\delta^2}{n\delta^2 + \sigma_o^2}\right).$$

Se deduce que el estimador de Bayes para μ es

$$\hat{\mu}_B = \frac{\delta^2 U + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2} = \frac{n\delta^2}{n\delta^2 + \sigma_o^2}\overline{Y} + \frac{\sigma_o^2}{n\delta^2 + \sigma_o^2}\eta.$$

De nuevo, este estimador de Bayes es un promedio ponderado de la UME, $\overline{Y}$, la media muestral y la media de la previa η. Cuando aumenta el tamaño de la muestra n, el valor asignado a la media muestral $\overline{Y}$ aumenta mientras que el valor asignado a la media previa η disminuye. ■

Ejercicios

16.1 Consulte los resultados del Ejemplo 16.2 dados en la Tabla 16.1.

- **a** ¿Cuál de las dos previas tiene la menor varianza?
- **b** Compare las medias y varianzas de las dos posteriores asociadas con la previa beta (1, 3). ¿Cuál de las posteriores tiene media y varianza que difiere más de la media y la varianza de la previa beta (1, 3)?
- **c** Conteste las preguntas de los incisos a y b para la previa beta (10, 30).
- **d** ¿Sus respuestas a los incisos a–c están apoyadas por las gráficas presentadas en la Figura 16.1(a) y (b)?
- **e** Compare las posteriores con base en $n = 5$ para las dos previas. ¿Cuál de las dos posteriores tiene media y varianza que difiere más de la media y varianza de las previas correspondientes?

16.2 Defina lo siguiente:

- **a** Distribución previa para un parámetro θ.
- **b** Distribución posterior para un parámetro θ.
- **c** Distribución previa conjugada.
- **d** Estimador de Bayes para una función de θ, $t(\theta)$

16.3 **Ejercicio Applet** La aplicación *Binomial Revision* se puede usar para explorar el impacto de los datos y la previa sobre la distribución posterior del parámetro p de Bernoulli. La demostración en la parte superior de la pantalla utiliza la previa beta con $\alpha = \beta = 1$.

a Haga clic en el botón "Next Trial" para observar el resultado de tomar una muestra de tamaño $n = 1$ de una población de Bernoulli con $p = .4$. ¿Observó un éxito o un fracaso? ¿La posterior se ve diferente a la previa? ¿Los parámetros de la posterior son lo que se esperaba con base en los resultados teóricos del Ejemplo 16.1?

b Haga clic en el botón "Next Trial" una vez más para observar el resultado y tomar una muestra de tamaño total $n = 2$ de una población de Bernoulli con $p = .4$. ¿Cuántos éxitos y fracasos ha observado usted hasta ahora? ¿La posterior se ve diferente de la posterior que usted obtuvo en el inciso a? ¿Los parámetros de la posterior son lo que usted esperaba con base en los resultados teóricos del Ejemplo 16.1?

c Haga clic en el botón "Next Trial" varias veces para observar el resultado de tomar muestras de tamaños más grandes de una población de Bernoulli con $p = .4$. Ponga atención a la media y la varianza de la distribución posterior que obtuvo al tomar muestras sucesivamente más grandes. ¿Qué observa acerca de los valores de las medias de las posteriores? ¿Qué observa acerca de las desviaciones estándar de las posteriores basadas en tamaños muestrales más grandes?

d En la demostración inicial de la aplicación se le comunicó que el verdadero valor del parámetro de Bernoulli es $p = .4$. La media de la previa beta con $\alpha = \beta = 1$ es .5. ¿Cuántos intentos son necesarios para obtener una posterior con media cercana a .4, el verdadero valor del parámetro de Bernoulli?

e Haga clic en el botón "50 Trials" para ver el efecto de los resultados de 50 intentos adicionales a la posterior. ¿Qué observa usted acerca de la forma de las distribuciones posteriores con base en un número grande de intentos?

16.4 **Ejercicio Applet** Arrastre el cursor a la sección "Applet with Controls" en la aplicación *Binomial Revision.* Aquí, el usuario puede establecer el verdadero valor del parámetro p de Bernoulli para cualquier valor $0 < p < 1$ (cualquier valor de interés "real") y puede seleccionar cualquier $\alpha > 0$ y $\beta > 0$ como los valores de los parámetros de la previa beta conjugada. ¿Qué pasará si el verdadero valor de $p = .1$ y se selecciona una previa beta con media $1/4$? En el Ejemplo 16.1, uno de estos conjuntos de valores para α y β se ilustró: $\alpha = 1$, $\beta = 3$. Inicie la aplicación para simular muestreo de una distribución de Bernoulli con $p = .1$ y use la previa beta (1, 3). (Asegúrese de presionar Enter después de introducir los valores apropiados en las cajas.)

a Haga clic en el botón "Next Trial" para observar el resultado de tomar una muestra de tamaño $n = 1$ de una población de Bernoulli con $p = .1$. ¿Observó un éxito o un fracaso? ¿La posterior se ve diferente a la previa?

b Haga clic en el botón "Next Trial" una vez más para observar el resultado y tomar una muestra de tamaño total $n = 2$ de una población de Bernoulli con $p = .1$. ¿Cuántos éxitos y fracasos ha observado hasta ahora? ¿La posterior se ve diferente de la posterior que usted obtuvo en el inciso a?

c Si observó un éxito en cualquiera de los dos primeros intentos, haga clic en el botón "Reset" y empiece de nuevo. A continuación, haga clic en el botón "Next Trial" hasta que observe el primer éxito. ¿Qué le ocurre a la forma de la posterior al observar el primer éxito?

d En esta demostración supusimos que el verdadero valor del parámetro de Bernoulli es $p = .1$. La media de la beta previa con $\alpha = 1$, $\beta = 3$ es .25. Haga clic en el botón "Next Trial" hasta que obtenga una posterior que tenga media cercana a .1. ¿Cuántos intentos son necesarios?

16.5 Repita las instrucciones del Ejercicio 16.4, usando una beta previa con $\alpha = 10$, $\beta = 30$. ¿Cómo se compara el número de intentos necesarios para obtener una posterior con media cercana a .1 con el número que el lector encontró en el Ejercicio 16.4(d)?

16.6 Suponga que Y es una variable aleatoria binomial basada en n intentos y probabilidad de éxito p (éste es el caso para el ejemplo de una enfermedad virulenta de la Sección 16.1). Use la beta previa conjugada con parámetros α y β para deducir la distribución posterior de $p \mid y$. Compare esta posterior con la hallada en el Ejemplo 16.1.

16.7 En la Sección 16.1 y el Ejercicio 16.6, consideramos un ejemplo donde el número de quienes responden a un tratamiento para una enfermedad virulenta en una muestra de tamaño n, tuvieron una distribución binomial con parámetro p y se usó una beta previa para p con parámetros $\alpha = 1$ y $\beta = 3$.

a Encuentre el estimador de Bayes para $p =$ la proporción de quienes tienen la enfermedad virulenta que respondieron a la terapia.

b Deduzca la media y la varianza del estimador de Bayes halladas en el inciso a.

16.8 Consulte el Ejercicio 16.6. Si Y es una variable aleatoria binomial basada en n intentos y probabilidad p de éxito y p tiene la beta previa conjugada con parámetros $\alpha = 1$ y $\beta = 1$,

a determine el estimador de Bayes para p, $\hat{p}_B$.

b ¿cuál es el otro nombre para la distribución beta con $\alpha = 1$ y $\beta = 1$?,

c encuentre la media cuadrática de error (MSE) del estimador de Bayes hallada en el inciso a. [*Sugerencia:* recuerde el Ejercicio 8.17].

d ¿Para qué valores de p es la media cuadrática de error del estimador de Bayes menor que la del estimador insesgado $\hat{p} = Y/n$?

16.9 Suponga que realizamos intentos de Bernoulli independientes y registramos Y, el número del intento en el que ocurre el primer éxito. Como ya vimos en la Sección 3.5, la variable aleatoria Y tiene una distribución geométrica con probabilidad p de éxito. Una distribución beta es de nuevo una previa conjugada para p.

a Si escogemos una previa beta con parámetros α y β, demuestre que la distribución posterior de $p \mid y$ es beta con parámetros $\alpha^\star = \alpha + 1$ y $\beta^\star = \beta + y - 1$.

b Encuentre los estimadores de Bayes para p y $p(1 - p)$.

16.10 Denote con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población exponencialmente distribuida con densidad $f(y \mid \theta) = \theta e^{-\theta y}, 0 < y$. (*Nota:* la media de esta población es $\mu = 1/\theta$). Use la previa conjugada gamma (α, β) para θ para hacer lo siguiente.

a Demostrar que la densidad conjunta de $Y_1, Y_2, \ldots, Y_n, \theta$ es

$$f(y_1, y_2, \ldots, y_n, \theta) = \frac{\theta^{n+\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left[-\theta \Big/ \left(\frac{\beta}{\beta \sum y_i + 1}\right)\right].$$

b Demostrar que la densidad marginal de $Y_1, Y_2, \ldots, Y_n$ es

$$m(y_1, y_2, \ldots, y_n) = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\beta}{\beta \sum y_i + 1}\right)^{\alpha+n}.$$

c Demostrar que la densidad posterior para $\theta|(y_1, y_2, \ldots, y_n)$ es una densidad gamma con parámetros $\alpha^\star = n + \alpha$ y $\beta^\star = \beta/\left(\beta \sum y_i + 1\right)$.

d Demostrar que el estimador de Bayes para $\mu = 1/\theta$ es

$$\hat{\mu}_B = \frac{\sum Y_i}{n + \alpha - 1} + \frac{1}{\beta(n + \alpha - 1)}.$$

[*Sugerencia:* recuerde el Ejercicio 4.111(e).]

e Demostrar que el estimador de Bayes del inciso d se puede escribir como un promedio ponderado de $\overline{Y}$ y la media previa para $1/\theta$. [*Sugerencia:* recuerde el Ejercicio 4.111(e).]

f Demostrar que el estimador de Bayes del inciso d es un estimador sesgado pero consistente para $\mu = 1/\theta$.

16.11 Denote con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población con distribución de Poisson con media λ. En este caso, $U = \sum Y_i$ es un estadístico suficiente para λ, y U tiene una distribución de Poisson con media $n\lambda$. Use la previa conjugada de gamma (α, β) para λ para hacer lo siguiente.

a Demostrar que la verosimilitud conjunta de U, λ es

$$L(u, \lambda) = \frac{n^u}{u!\beta^\alpha \Gamma(\alpha)} \lambda^{u+\alpha-1} \exp\left[-\lambda \Big/ \left(\frac{\beta}{n\beta + 1}\right)\right].$$

b Demostrar que la función marginal de masa de U es

$$m(u) = \frac{n^u \Gamma(u + \alpha)}{u!\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{n\beta + 1}\right)^{u+\alpha}.$$

c Demostrar que la densidad posterior para $\lambda \mid u$ es una densidad gamma con parámetros $\alpha^* = u + \alpha$ y $\beta^* = \beta/(n\beta + 1)$.

d Demostrar que el estimador de Bayes para λ es

$$\hat{\lambda}_B = \frac{\left(\sum Y_i + \alpha\right)\beta}{n\beta + 1}.$$

e Demostrar que el estimador de Bayes del inciso d se puede escribir como un promedio ponderado de $\overline{Y}$ y la media previa para λ.

f Demostrar que el estimador de Bayes del inciso d es un estimador sesgado pero consistente para λ.

16.12 Denote con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una población normal con una media *conocida* μ_o y varianza desconocida $1/\nu$. En este caso, $U = \sum(Y_i - \mu_o)^2$ es un estadístico suficiente para ν, y $W = \nu U$ tiene una distribución χ^2 con n grados de libertad. Use la previa conjugada gamma (α, β) para ν para hacer lo siguiente.

a Demostrar que la densidad conjunta de U, ν es

$$f(u, \nu) = \frac{u^{(n/2)-1} \nu^{(n/2)+\alpha-1}}{\Gamma(\alpha)\Gamma(n/2)\beta^\alpha 2^{(n/2)}} \exp\left[-\nu \Big/ \left(\frac{2\beta}{u\beta + 2}\right)\right].$$

b Demostrar que la densidad marginal de U es

$$m(u) = \frac{u^{(n/2)-1}}{\Gamma(\alpha)\Gamma(n/2)\beta^\alpha 2^{(n/2)}} \left(\frac{2\beta}{u\beta + 2}\right)^{(n/2)+\alpha} \Gamma\left(\frac{n}{2} + \alpha\right).$$

c Demostrar que la densidad posterior para $\nu|u$ es una densidad gamma con parámetros $\alpha^* = (n/2) + \alpha$ y $\beta^* = 2\beta/(u\beta + 2)$.

d Demostrar que el estimador de Bayes para $\sigma^2 = 1/\nu$ es $\hat{\sigma}^2_B = (U\beta + 2)/[\beta(n + 2\alpha - 2)]$. [*Sugerencia:* recuerde el Ejercicio 4.111(e).]

e La MLE para σ^2 es U/n. Demuestre que el estimador de Bayes del inciso d se puede escribir como un promedio ponderado de la MLE y la media previa de $1/\nu$. [*Sugerencia:* recuerde el Ejercicio 4.111(e).]

16.3 Intervalos creíbles de Bayes

En secciones previas hemos determinado cómo deducir intervalos de confianza clásicos para varios parámetros de interés. En nuestra proposición previa, el parámetro de interés θ tenía un *valor fijo pero desconocido*. Construimos intervalos al hallar dos *variables aleatorias* $\hat{\theta}_L$ y $\hat{\theta}_U$, los límites de confianza inferior y superior, tales que $\hat{\theta}_L < \hat{\theta}_U$ y de modo que la probabilidad de que el *intervalo aleatorio* $(\hat{\theta}_L, \hat{\theta}_U)$ encerrara el valor *fijo* de θ era igual al coeficiente de confianza prescrito $1 - \alpha$. También consideramos cómo formar regiones de confianza unilaterales. La realización clave en nuestro trabajo antes de Bayes fue que el *intervalo* era aleatorio y el parámetro era fijo. En el Ejemplo 8.11 construimos un intervalo de confianza para la media de una población normalmente distribuida con varianza desconocida usando la fórmula

$$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) = \left\{\overline{Y} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right),\ \overline{Y} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right\}.$$

En este caso los puntos extremos superior e inferior del intervalo son claramente variables aleatorias. Al obtener los datos, calculando los valores *realizados* de la *media muestral* $\overline{y} = 2959$ y la varianza muestral $s = 39.1$ y usando $n = 8$ y $t_{.025} = 2.365$, determinamos que nuestro intervalo de confianza *realizado* para la velocidad media en la boca del cañón para granadas del tipo considerado es (2926.3, 2991.7). Éste es un intervalo fijo que contiene la verdadera velocidad media en la boca del cañón o no la contiene. Decimos que es un intervalo de confianza de 95% porque *si se tomaran muestras independientes y separadas, cada una de tamaño* $n = 8$ y se determinaran los intervalos resultantes (diferentes), a la larga, 95% *de los intervalos contendrían la verdadera media.* El parámetro es fijo, los puntos extremos del intervalo son aleatorios y diferentes muestras darán intervalos realizados diferentes.

En el contexto de Bayes, el parámetro θ es una *variable aleatoria* con función de densidad posterior $g^\star(\theta)$. Si consideramos el intervalo (a, b), la probabilidad posterior de que la variable aleatoria θ se encuentre en este intervalo es

$$P^\star(a \le \theta \le b) = \int_a^b g^\star(\theta)\, d\theta.$$

Si la probabilidad posterior $P^\star(a \le \theta \le b) = .90$, decimos que (a, b) es un *intervalo creíble* de 90% para θ.

EJEMPLO **16.5** En el Ejemplo 8.11 fue razonable suponer que las velocidades en la boca del cañón estaban normalmente distribuidas con media desconocida μ. En ese ejemplo supusimos que la varianza de las velocidades en la boca del cañón σ^2 era desconocida. Suponga ahora que estamos interesados en formar un intervalo creíble de Bayes para μ y suponemos que hay una alta probabilidad de que las velocidades en la boca del cañón estarán a no más de 30 pies por segundo de la media μ. Debido a que una población normalmente distribuida es tal que aproximadamente 95% de sus valores están a no más de 2 desviaciones estándar de su media, podría ser razonable suponer que la distribución fundamental de velocidades en la boca del cañón está normalmente distribuida con media μ y varianza σ_o^2 tal que $2\sigma_o = 30$, que está con $\sigma_o^2 = 225$.

Si antes de observar cualquier información pensamos que hay una alta probabilidad de que μ esté entre 2700 y 2900, podríamos escoger usar una previa normal conjugada para μ con media η y varianza δ^2 escogidas de modo que $\eta - 2\delta = 2700$ y $\eta + 2\delta = 2900$ o $\eta = 2800$ y $\delta^2 = 50^2 = 2500$. Observe que hemos supuesto considerablemente más conocimiento de las velocidades en la boca del cañón que en el Ejemplo 8.11, donde sólo supusimos que las velocidades en la boca del cañón estaban normalmente distribuidas (con varianza desconocida). Si estamos más cómodos con esta estructura adicional, ahora tomamos nuestra muestra de tamaño $n = 8$ y obtenemos las velocidades en la boca del cañón dadas a continuación:

3005	2925	2935	2965
2995	3005	2937	2905

Use la forma general para la densidad posterior para $\mu \mid u$ desarrollada en el Ejemplo 16.4 para dar un intervalo creíble de 95% para μ.

Solución Esta situación es un caso especial del que hablamos en el Ejemplo 16.4. En esta aplicación de ese resultado general,

$$n = 8, \qquad u = \sum y_i = 23\,672, \qquad \sigma_o^2 = 225, \qquad \eta = 2800, \qquad \delta^2 = 2500.$$

En el Ejemplo 16.4 determinamos que la densidad posterior de $\mu \mid u$ es una densidad normal con media η^* y varianza δ^{*2} dadas por

$$\eta^\star = \frac{\delta^2 u + \sigma_o^2 \eta}{n\delta^2 + \sigma_o^2} = \frac{(2500)\,(23\,672) + (225)(2800)}{8(2500) + 225} = 2957.23,$$

$$\delta^{\star 2} = \frac{\sigma_o^2 \delta^2}{n\delta^2 + \sigma_o^2} = \frac{(225)(2500)}{8(2500) + 225} = 27.81.$$

Finalmente, recuerde que cualquier variable aleatoria normalmente distribuida W con media μ_W y varianza σ_W^2 es tal que

$$P(\mu_W - 1.96\,\sigma_W \le W \le \mu_W + 1.96\,\sigma_W) = .95.$$

Se deduce que un intervalo creíble de 95% para μ es

$$\begin{aligned}(\eta^* - 1.96\,\delta^*, \eta^* + 1.96\,\delta^*) &= (2957.23 - 1.96\sqrt{27.81},\ 2957.23 + 1.96\sqrt{27.81}) \\ &= (2946.89,\ 2967.57).\end{aligned}$$

■

Es importante observar que diferentes personas que construyan intervalos creíbles para μ, usando los datos del Ejemplo 16.5, obtendrán otros intervalos si escogen distintos valores para cualquiera de los parámetros η, δ^2 y σ_o^2. No obstante, para las opciones empleadas en el Ejemplo 16.5, al combinar su conocimiento previo con la información de los datos, el analista puede decir que la probabilidad posterior es .95 de que μ (aleatoria) está en el intervalo (fijo) de (2946.89, 2967.57).

EJEMPLO **16.6** En el Ejercicio 16.10 dijimos que si $Y_1, Y_2, \ldots, Y_n$ denotan una muestra aleatoria de una población exponencialmente distribuida con densidad $f(y \mid \theta) = \theta e^{-\theta y}$, $0 < y$, y se utilizó la previa conjugada gamma (con parámetros α y β) para θ, entonces la densidad posterior para θ es una densidad gamma con parámetros $\alpha^* = n + \alpha$ y $\beta^* = \beta/(\beta \sum y_i + 1)$. Suponga que un analista escogió $\alpha = 3$ y $\beta = 5$ como valores de parámetro apropiados para la previa y que una muestra de tamaño $n = 10$ dio que $\sum y_i = 1.26$. Construya intervalos creíbles de 90% para θ y la media de la población, $\mu = 1/\theta$.

Solución En esta aplicación del resultado general dado en el Ejercicio 16.10,

$$n = 10, \qquad u = \sum y_i = 1.26, \qquad \alpha = 3, \qquad \beta = 5.$$

La densidad posterior resultante de θ es una densidad gamma con α^* y β^* dadas por

$$\alpha^* = n + \alpha = 10 + 3 = 13,$$

$$\beta^* = \frac{\beta}{\beta \sum y_i + 1} = \frac{5}{5(1.26) + 1} = .685.$$

Para completar nuestros cálculos, necesitamos hallar dos valores a y b tales que

$$P^{\star}(a \le \theta \le b) = .90.$$

Si lo hacemos así, un intervalo creíble de 90% para θ es (a, b). Además, como

$$a \le \theta \le b \text{ si y sólo si } 1/b \le 1/\theta \le 1/a,$$

se deduce que un intervalo creíble de 90% para $\mu = 1/\theta$ es $(1/b, 1/a)$.

Aun cuando no tenemos una tabla para obtener probabilidades asociadas con variables aleatorias de distribución gamma con valores de parámetro diferentes, estas probabilidades se pueden hallar usando una de las aplicaciones breves disponibles en www.thomsonedu.com/statistics/wackerly. *R, S*-Plus y otros programas de software de estadística se pueden usar también para calcular probabilidades asociadas con variables de distribución gamma. Incluso así, habrá un número infinito de opciones para a y b tales que $P^{*}(a \le \theta \le b) = .90$. Si hallamos valores de a y b tales que

$$P^{\star}(\theta \ge a) = .95 \quad \text{y} \quad P^{\star}(\theta \ge b) = .05,$$

estos valores necesariamente satisfacen nuestro requisito inicial de que $P^{*}(a \le \theta \le b) = .90$.

En nuestra aplicación presente determinamos que θ tiene una posterior gamma con parámetros $\alpha^{*} = 13$ y $\beta^{*} = .685$. Usando la aplicación *Gamma Probabilities and Quantiles* en el sitio web de Thomson, determinamos que

$$P^{*}(\theta \ge 5.2674) = .95 \quad \text{y} \quad P^{*}(\theta \ge 13.3182) = .05.$$

Entonces, para los datos observados y la previa que seleccionamos, (5.2674, 13.3182) es un intervalo creíble de 90% para θ mientras que $[1/(13.3182), (1/5.7624)] = (.0751, .1898)$ es un intervalo creíble de 90% para $\mu = 1/\theta$.

El comando `qgamma(.05,13,1/.685)` de *R* (o *S*-plus) también da el valor $a = 5.2674$ dado líneas antes, mientras que `qgamma(.95,13,1/.685)` da $b = 13.3182$. ■

Ejercicios

16.13 **Ejercicio Applet** Active la aplicación *Binomial Revision* y arrastre el cursor a la sección marcada "Credible Interval". Cambie el valor de la proporción de Bernoulli a 0.45 y los parámetros de la previa beta a $\alpha = 3$ y $\beta = 5$ y presione Enter en su computadora.

a ¿Cuál es el intervalo creíble sin datos para p con base en la previa beta (3, 5)?

b Use la aplicación *Beta Probabilities and Quantiles* (accesible en www.thomsonedu.com/statistics/wackerly) para calcular la probabilidad previa de que p es mayor que el punto extremo superior del intervalo que usted obtuvo en el inciso a. También calcule la probabilidad de que p sea menor que el punto extremo inferior del intervalo que obtuvo en el inciso a.

c Con base en sus respuestas al inciso b, ¿cuál es la probabilidad previa de que p esté en el intervalo que obtuvo en el inciso a? ¿Está usted de acuerdo en que el intervalo obtenido en el inciso a es un intervalo creíble de 95% para p con base en la previa beta (3, 5)?

d Haga clic en el botón "Next Trial" una vez. ¿La posterior basada en la muestra de tamaño 1 es diferente a la previa? ¿Cómo difiere la posterior de la previa?

e ¿Cuál es un intervalo creíble de 95% basado en la previa y el resultado de su muestra de tamaño 1? ¿Es mayor o menor que el intervalo que obtuvo (sin datos) en el inciso a?

f Haga clic en el botón "Next Trial" una vez más. Compare la longitud de este intervalo (con base en los resultados de una muestra de tamaño 2) con los intervalos obtenidos en los incisos a y e.

g Use la aplicación *Beta Probabilities and Quantiles* para calcular la probabilidad posterior de que p sea mayor que el punto extremo superior del intervalo que obtuvo en el inciso f. ¿El valor de esta probabilidad posterior le sorprende?

h Haga clic en el botón "Next Trial" varias veces. Describa la forma en que la posterior cambia con datos adicionales. ¿Qué observa usted acerca de las longitudes de los intervalos creíbles obtenidos usando posteriores con base en tamaños muestrales más grandes?

16.14 **Ejercicio Applet** Consulte el Ejercicio 16.13. Seleccione un valor para el verdadero valor de la proporción p de Bernoulli y valores para los parámetros de la previa beta conjugada.

a Repita el Ejercicio 16.13 a–h, usando los valores que seleccionó.

b Haga clic también en el botón "50 Trials" unas pocas veces. Observe los valores de las sucesivas desviaciones estándar posteriores y las longitudes de los intervalos creíbles sucesivos.

i ¿Qué observa acerca de las desviaciones estándar de las distribuciones posteriores sucesivas?

ii Con base en su respuesta al inciso i, ¿qué efecto espera observar acerca de las longitudes de intervalos creíbles sucesivos?

iii ¿Las longitudes de los intervalos creíbles sucesivos se comportaron como usted anticipó en el inciso ii?

16.15 **Ejercicio Applet** En el Ejercicio 16.7 reconsideramos nuestro ejemplo de introducción donde el número de quienes responden a un tratamiento para una enfermedad virulenta, en una muestra de tamaño n, tuvo una distribución binomial con parámetro p y usamos una beta previa para p con parámetros $\alpha = 1$ y $\beta = 3$. Posteriormente hallamos que, al observar $Y = y$ que respondieron, la función de densidad posterior para $p \mid y$ es una densidad beta con parámetros $\alpha^* = y + \alpha = y + 1$ y $\beta^* = n - y + \beta = n - y + 3$. Si obtuvimos una muestra de tamaño $n = 25$ que contenía 4 personas que respondieron al nuevo tratamiento, encuentre un intervalo creíble de 95% para p. [Utilice la aplicación *Beta Probabilities and Quantiles* en www.thomsonedu.com/statistics/wackerly. Alternativamente, si W es una variable aleatoria de distribución beta con parámetros α y β, el comando `qbeta(p,α,β)` de *R* (o *S*-Plus) da el valor w tal que $P(W \le w) = p$.]

16.16 **Ejercicio Applet** Repita las instrucciones para el Ejercicio 16.15 suponiendo una beta previa con parámetros $\alpha = 1$ y $\beta = 1$ [una previa que es uniforme en el intervalo (0, 1)]. (Vea el resultado del Ejercicio 16.8.) Compare este intervalo con el obtenido en el Ejercicio 16.15.

16.17 **Ejercicio Applet** En el Ejercicio 16.9 empleamos una beta previa con parámetros α y β y hallamos la densidad posterior para el parámetro p asociada con una distribución geométrica. Determinamos que la distribución posterior de $p \mid y$ es beta con parámetros $\alpha^* = \alpha + 1$ y $\beta^* = \beta + y - 1$. Suponga que usamos $\alpha = 10$ y $\beta = 5$ en nuestra beta previa y observamos el primer éxito en el intento 6. Determine un intervalo creíble de 80% para p.

16.18 **Ejercicio Applet** En el Ejercicio 16.10 hallamos la densidad posterior para θ con base en una muestra de tamaño n de una población exponencialmente distribuida con media $1/\theta$. Específicamente, usando la densidad gamma con parámetros α y β como la previa para θ, hallamos que la densidad posterior para $\theta \mid (y_1, y_2, \ldots, y_n)$ es una densidad gamma con parámetros $\alpha^* = n + \alpha$ y $\beta^* = \beta / \left(\beta \sum y_i + 1\right)$. Suponiendo que una muestra de tamaño $n = 15$ produjo una muestra tal que $\sum y_i = 30.27$ y que los parámetros de la gamma previa son $\alpha = 2.3$ y $\beta = 0.4$, use la aplicación *Gamma Probabilities and Quantiles* para hallar intervalos creíbles de 80% para θ y $1/\theta$, la media de la población exponencial.

16.19 **Ejercicio Applet** En el Ejercicio 16.11 encontramos la densidad posterior para λ, la media de una población con distribución de Poisson. Suponiendo una muestra de tamaño n y una gamma previa conjugada (α, β) para λ, demostramos que la densidad posterior de $\lambda \mid \sum y_i$ es gamma con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = \beta/(n\beta + 1)$. Si una muestra de tamaño $n = 25$ es tal que $\sum y_i = 174$ y los parámetros previos fueron $(\alpha = 2, \beta = 3)$, use la aplicación *Gamma Probabilities and Quantiles* para hallar un intervalo creíble de 95% para λ.

16.20 **Ejercicio Applet** En el Ejercicio 16.12 empleamos una gamma previa (α, β) para ν y una muestra de tamaño n de una población normal con media μ_o conocida y varianza $1/\nu$ para deducir la posterior para ν. Específicamente, si $u = \sum(y_i - \mu_o)^2$, determinamos que la posterior de $\nu|u$ es gamma con parámetros $\alpha^* = (n/2) + \alpha$ y $\beta^* = 2\beta/(u\beta + 2)$. Si escogemos que los parámetros de la previa sean $(\alpha = 5, \beta = 2)$ y una muestra de tamaño $n = 8$ da el valor $u = .8579$, use la aplicación *Gamma Probabilities and Quantiles* para determinar intervalos creíbles de 90% para ν y $1/\nu$, la varianza de la población de la cual se obtuvo la muestra.

16.4 Pruebas de hipótesis de Bayes

Las pruebas de hipótesis también se pueden abordar desde una perspectiva de Bayes. Como hemos visto en secciones previas, el método de Bayes emplea información previa acerca de un parámetro *e* información en los datos acerca de ese parámetro para obtener la distribución posterior. Si, como en la Sección 10.11 en la que consideramos pruebas de razón de verosimilitud, estamos interesados en probar que el parámetro θ se encuentra en uno de dos conjuntos de valores, Ω_0 y Ω_a, podemos usar la distribución posterior de θ para calcular la probabilidad posterior de que θ esté en cada uno de estos conjuntos de valores. Cuando se pruebe $H_0 : \theta \in \Omega_0$ contra $H_a : \theta \in \Omega_a$, un método que se usa con frecuencia es calcular las probabilidades posteriores $P^*(\theta \in \Omega_0)$ y $P^*(\theta \in \Omega_a)$ y aceptar la hipótesis con la más alta probabilidad posterior. Esto es, para probar $H_0 : \theta \in \Omega_0$ contra $H_a : \theta \in \Omega_a$,

aceptar H_o si $P^\star(\theta \in \Omega_0) > P^\star(\theta \in \Omega_a)$,
aceptar H_a si $P^\star(\theta \in \Omega_a) > P^\star(\theta \in \Omega_0)$.

EJEMPLO **16.7** En el Ejemplo 16.5 obtuvimos un intervalo creíble de 95% para la velocidad media en la boca del cañón asociada con granadas elaboradas con pólvora reformulada. Supusimos que las velocidades asociadas en la boca del cañón están normalmente distribuidas con media μ y varianza $\sigma_o^2 = 225$ y que una densidad previa razonable para μ es normal con media $\eta = 2800$ y varianza $\delta^2 = 2500$. Entonces usamos los datos

3005	2925	2935	2965
2995	3005	2937	2905

para obtener que la densidad posterior para μ es normal con media $\eta^\star = 2957.23$ y desviación estándar $\delta^\star = 5.274$. Efectúe la prueba de Bayes para

$$H_0 : \mu \leq 2950 \quad \text{contra} \quad H_a : \mu > 2950.$$

Solución En este caso, si Z tiene una distribución normal estándar,

$$
\begin{aligned}
P^\star(\theta \in \Omega_0) &= P^\star(\mu \leq 2950) \\
&= P\left(Z \leq \frac{2950 - \eta^\star}{\delta^\star}\right) = P\left(Z \leq \frac{2950 - 2957.23}{5.274}\right) \\
&= P(Z \leq -1.37) = .0951,
\end{aligned}
$$

y $P^\star(\theta \in \Omega_a) = P^\star(\mu > 2950) = 1 - P^\star(\mu \leq 2950) = .9049$. Entonces, vemos que la probabilidad posterior de H_a es mucho mayor que la probabilidad posterior de H_0 y nuestra decisión es aceptar H_a: $\mu > 2950$. ■

De nuevo, observamos que si un analista diferente emplea los mismos datos para efectuar una prueba de Bayes para las mismas hipótesis pero con diferentes valores para cualquier η, δ^2 y σ_o^2, obtendrá probabilidades posteriores de las hipótesis que son diferentes a las obtenidas en el Ejemplo 16.7. Entonces, diferentes analistas con distintas opciones de valores para los parámetros previos podrían llegar a conclusiones diferentes.

En las situaciones más frecuentes examinadas en los capítulos previos, el parámetro θ tiene un valor fijo pero desconocido y cualquier hipótesis es verdadera o falsa. Si $\theta \in \Omega_0$, entonces la hipótesis nula es ciertamente verdadera (con probabilidad 1), y la alternativa es ciertamente falsa. Si $\theta \in \Omega_a$, entonces la hipótesis alternativa es ciertamente verdadera (con probabilidad 1) y la hipótesis nula es ciertamente falsa. La única forma en que podríamos *saber* si $\theta \in \Omega_0$ es *si conocemos el verdadero valor de* θ. Si éste fuera el caso, efectuar una prueba de hipótesis sería superfluo. Por esta razón, la prueba más frecuente no hace referencia a las probabilidades de las hipótesis sino que se concentra en la probabilidad de un error tipo I, α, y el poder de la prueba, poder $(\theta) = 1 - \beta(\theta)$. Por el contrario, los conceptos más frecuentes de tamaño y poder no son de interés para un analista que utilice una prueba de Bayes.

EJEMPLO 16.8 En el Ejemplo 16.6 utilizamos un resultado dado en el Ejercicio 16.7 para obtener intervalos creíbles para θ y la media poblacional μ basada en $Y_1, Y_2, \ldots, Y_n$, una muestra aleatoria de una población exponencialmente distribuida con densidad $f(y|\theta) = \theta e^{-\theta y}$, $0 < y$. Usando una previa gamma conjugada por θ con parámetros $\alpha = 3$ y $\beta = 5$ obtenemos que la densidad posterior para θ es una densidad gamma con parámetros $\alpha^* = 13$ y $\beta^* = .685$. Efectúe una prueba de Bayes para

$$H_0: \mu > .12 \quad \text{contra} \quad H_a: \mu \leq .12.$$

Solución Como la media de la distribución exponencial es $\mu = 1/\theta$, las hipótesis son equivalentes a

$$H_0: \theta < 1/(.12) = 8.333 \quad \text{contra} \quad H_a: \theta \geq 8.333$$

Como la densidad posterior para θ es una densidad gamma con parámetros $\alpha^* = 13$ y $\beta^* = .685$,

$$P^*(\theta \in \Omega_0) = P^*(\theta < 8.333) \quad \text{y} \quad P^*(\theta \in \Omega_a) = P^*(\theta \geq 8.333).$$

En nuestra aplicación presente, determinamos que θ tiene una posterior gamma con parámetros $\alpha^* = 13$ y $\beta^* = .685$. Usando la aplicación *Gamma Probabilities and Quantiles*,

$$P^*(\theta \in \Omega_a) = P^*(\theta \geq 8.333) = 0.5570,$$

y

$$P^*(\theta \in \Omega_0) = P^*(\theta < 8.333) = 1 - P^*(\theta \geq 8.333) = 0.4430.$$

En este caso, la probabilidad posterior de H_a es un poco mayor que la probabilidad posterior de H_0. Es decisión del analista decidir si las probabilidades son suficientemente diferentes para ameritar la decisión de aceptar $H_a : \mu \leq .12$.

Si usted prefiere usar *R* o *S*-Plus para calcular las probabilidades posteriores de las hipótesis, `pgamma(8.333,13,1/.685)` da $P^*(\theta \in \Omega_0) = P^*(\theta < 8.333)$ y $P^*(\theta \in \Omega_a) = P^*(\theta \geq 8.333) = 1 - P^*(\theta \in \Omega_0)$. ■

Ejercicios

16.21 **Ejercicio Applet** En el Ejercicio 16.15 determinamos que la densidad posterior para p, la proporción de quienes respondieron al nuevo tratamiento para una enfermedad virulenta, es una densidad beta con parámetros $\alpha^* = 5$ y $\beta^* = 24$. ¿Cuál es la conclusión de una prueba de Bayes para $H_0 : p < .3$ contra $H_a : p \geq .3$? [Use la aplicación *Beta Probabilities and Quantiles* en www.thomsonedu.com/statistics/wackerly. Alternativamente, si W es una variable aleatoria con distribución beta y parámetros α y β, el comando `pbeta(w,`α`,`β`)` de *R* o *S*-Plus da $P(W \leq w)$.]

16.22 **Ejercicio Applet** El Ejercicio 16.16 utilizó diferentes parámetros previos pero los mismos datos para determinar que la densidad posterior para p, la proporción de quienes respondieron al nuevo tratamiento para una enfermedad virulenta, es una densidad beta con parámetros $\alpha^* = 5$ y $\beta^* = 22$. ¿Cuál es la conclusión de una prueba de Bayes para $H_0 : p < .3$ contra $H_a : p \geq .3$? Compare su conclusión con la obtenida en el Ejercicio 16.21.

16.23 **Ejercicio Applet** En el Ejercicio 16.17 obtuvimos una posterior beta con parámetros $\alpha^* = 11$ y $\beta^* = 10$ para el parámetro p asociado con una distribución geométrica. ¿Cuál es la conclusión de una prueba de Bayes para $H_0 : p < .4$ contra $H_a : p \geq .4$?

16.24 **Ejercicio Applet** En el Ejercicio 16.18 hallamos que la densidad posterior para θ es una densidad gamma con parámetros $\alpha^* = 17.3$ y $\beta^* = .0305$. Dado que la media de la población exponencial fundamental es $\mu = 1/\theta$, probar las hipótesis $H_0 : \mu < 2$ contra $H_a : \mu \geq 2$ es equivalente a probar $H_0 : \theta > .5$ contra $H_a : \theta \leq .5$. ¿Cuál es la conclusión de una prueba de Bayes para estas hipótesis?

16.25 **Ejercicio Applet** En el Ejercicio 16.19 hallamos que la densidad posterior para λ, la media de una población con distribución de Poisson, es una densidad gamma con parámetros $\alpha^* = 176$ y $\beta^* = .0395$. ¿Cuál es la conclusión de una prueba de Bayes para $H_0 : \lambda > 6$ contra $H_a : \lambda \leq 6$?

16.26 **Ejercicio Applet** En el Ejercicio 16.20 determinamos que la posterior de $\nu \mid u$ es una densidad gamma con parámetros $\alpha^* = 9$ y $\beta^* = 1.0765$. Recuerde que $\nu = 1/\sigma^2$, donde σ^2 es la varianza de la población fundamental que está normalmente distribuida con media conocida μ_0. Probar las hipótesis $H_0 : \sigma^2 > 0.1$ contra $H_a : \sigma^2 \leq 0.1$ es equivalente a probar $H_0 : \nu < 10$ contra $H_a : \nu \geq 10$. ¿Cuál es la conclusión de una prueba de Bayes para estas hipótesis?

16.5 Resumen y comentarios adicionales

Como ya hemos visto en las secciones previas, la clave de los métodos inferenciales de Bayes (hallar estimadores, intervalos creíbles o realizar pruebas de hipótesis) es hallar la distribución posterior del parámetro θ. En especial cuando hay poca información, esta posterior depende en gran medida de la distribución fundamental y previa de la población de la cual se toma la muestra. Nos hemos enfocado en el uso de previas conjugadas debido a la sencillez resultante de hallar la distribución posterior pedida del parámetro de interés. Por supuesto que las previas conjugadas no son las únicas previas que se pueden usar, pero tienen la ventaja de resultar cálculos fáciles. Esto no significa que una previa conjugada sea necesariamente la elección correcta para la previa. Incluso si seleccionamos correctamente la familia de la cual se tome la previa (hemos hecho repetido uso de previas beta y gamma), prevalece la dificultad de seleccionar los valores apropiados asociados con los parámetros de la previa. Hemos visto, sin embargo, que la opción de los valores de parámetro para la previa tiene impacto decreciente para tamaños muestrales más grandes.

Quizá sea apropiado hacer algunos comentarios acerca de la selección de valores de los parámetros de la densidad previa. Si usamos una previa normal con media ν y varianza δ^2 y pensamos que es probable (o improbable) que el parámetro poblacional sea cercano a ν, usaríamos un valor relativamente pequeño (grande) para δ^2. Cuando se usa una previa beta con parámetros α y β para un parámetro que pensamos que tiene un valor cercano a c, podríamos seleccionar α y β tales que la media de la previa, $\alpha/(\alpha + \beta)$, es igual a c y la varianza de la previa, $\alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$, es pequeña. En el ejemplo de introducción empleamos una previa beta con $\alpha = 1$ y $\beta = 3$ porque pensamos que alrededor de 25% de quienes recibieron un nuevo tratamiento responderían favorablemente. La media y la desviación estándar de la posterior son, respectivamente, .25 y .1936. Observe que éstas no son las únicas opciones para α y β que den .25 como la media de la previa. En general, si $\alpha/(\alpha + \beta) = c$, entonces para cualquier $k > 0$, $\alpha' = k\alpha$ y $\beta' = k\beta$ también satisfacen $\alpha'/(\alpha' + \beta') = c$. No obstante, para una densidad beta con parámetros $\alpha' = k\alpha$ y $\beta' = k\beta$, la varianza de la previa es $\alpha'\beta'[(\alpha' + \beta')^2(\alpha' + \beta' + 1)] = \alpha\beta/[(\alpha + \beta)^2(k\alpha + k\beta + 1)]$. Por tanto, si nuestra selección inicial de α y β da un valor apropiado para la media de la previa pero preferimos una varianza más pequeña, podemos alcanzar esto al seleccionar alguna $k > 1$ y usar $\alpha' = k\alpha$ y $\beta' = k\beta$ como los parámetros previos. Por el contrario, al seleccionar alguna $k < 1$ y usar $\alpha' = k\alpha$ y $\beta' = k\beta$ como parámetros previos obtenemos la misma media previa pero varianza previa más grande. Por tanto, resulta una previa más vaga por seleccionar valores pequeños de α y β que sean tales que $\alpha/(\alpha + \beta) = c$, la media previa deseada.

Uno de los pasos al determinar la previa es definir la distribución marginal de los datos. Para previas continuas, esto se logra al integrar la verosimilitud conjunta de los datos y el parámetro sobre la región de apoyo para la previa. En nuestro trabajo previo denotamos la masa marginal resultante o función de densidad para las variables aleatorias $Y_1, Y_2, \ldots, Y_n$ en una muestra de tamaño n como $m(y_1, y_2, \ldots, y_n)$ o como $m(u)$ si U es un estadístico suficiente para θ. Esta función marginal de densidad o masa se denomina función *pronosticadora* de masa o densidad de los datos. Hemos dado explícitamente estas distribuciones pronosticadoras en todas nuestras aplicaciones. Esto se debe a que, parafraseando a Berger (1985, p. 95), el interés en la distribución pronosticadora se centra en el hecho de que ésta es la distribución de acuerdo con la cual la información ocurre en realidad. Como se afirma en Box (1980, pp. 385-386), la evidencia potencial de una selección inapropiada de modelo es proporcionada por la distribución pronosticadora de los datos, no por la distribución posterior para el parámetro. Algunos analistas expertos de Bayes escogen modelar la distribución pronosticadora directamente y seleccionar la previa que lleve a la distribución pronosticadora pedida. El reverendo Thomas Bayes (1784) utilizó una previa uniforme (0, 1) para el parámetro p de Bernoulli (o binomial) *porque* esta previa lleva a la distribución pronosticadora que él pensó era más apropiada. Comentarios adicionales relevantes para la selección de algunos parámetros previos se pueden hallar en la obra de Kepner y Wackerly (2002).

Sin oponerse al párrafo anterior, es cierto que hay un atajo para hallar la densidad posterior de suma importancia para θ. Como ya se indicó antes, si $L(y_1, y_2, \ldots, y_n | \theta)$ es la verosimilitud condicional de los datos y θ tiene densidad previa continua $g(\theta)$, entonces la densidad posterior de θ es

$$g^*(\theta \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}.$$

Observe que el denominador en el lado derecho de la expresión depende de $y_1, y_2, \ldots, y_n$, pero *no* depende de θ. (Definir una integración con respecto a θ produce un resultado que es libre de θ.) Al ver que con respecto a θ, el denominador es una constante, podemos escribir

$$g^*(\theta \mid y_1, y_2, \ldots, y_n) = c(y_1, y_2, \ldots, y_n) L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta),$$

donde

$$c(y_1, y_2, \ldots, y_n) = \frac{1}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}$$

no depende de θ. Además observe que, debido a que la densidad posterior es una función de densidad de buena fe, la cantidad $c(y_1, y_2, \ldots, y_n)$ *debe ser* tal que

$$\int_{-\infty}^{\infty} g^*(\theta \mid y_1, y_2, \ldots, y_n)\, d\theta = c(y_1, y_2, \ldots, y_n) \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta = 1.$$

Finalmente, vemos que la densidad posterior es *proporcional* al producto de la verosimilitud condicional de los datos y la densidad previa para θ:

$$g^*(\theta \mid y_1, y_2, \ldots, y_n) \propto L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta),$$

donde la constante de proporcionalidad se escoge de modo que la integral de la función de densidad posterior sea 1. Ilustramos esto al reconsiderar el Ejemplo 16.1.

EJEMPLO **16.9** Denotemos con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Bernoulli donde $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$ y supongamos que la distribución previa para p es beta (α, β). Encuentre la distribución posterior para p.

Solución Al igual que antes,

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid p)g(p) &= p(y_1, y_2, \ldots, y_n \mid p)g(p) \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i}\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}\right], \\ g^*(p \mid y_1, y_2, \ldots, y_n, p) &\propto p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}. \end{aligned}$$

De lo arriba expresado reconocemos que la resultante posterior para p debe ser beta con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = n - \sum y_i + \beta$. ■

¿Cuál fue la ventaja de hallar la posterior previa usando este argumento de "proporcionalidad"? ¡Considerablemente menos trabajo! ¿La desventaja? Nunca mostramos la función de masa pronosticadora para los datos y perdimos la oportunidad de criticar el modelo de Bayes.

Las previas que no sean conjugadas bien podrían ser más apropiadas en aplicaciones específicas. La posterior se encuentra usando el mismo procedimiento dado en la Sección 16.2, pero podríamos obtener una distribución posterior con la cual no estemos familiarizados. Hallar la media de la posterior, intervalos creíbles y las probabilidades de hipótesis relevantes podría ser más problemático. Para los ejemplos de las secciones previas obtuvimos posteriores que conocíamos muy bien. Las medias posteriores fueron fáciles de hallar porque ya habíamos determinado propiedades de variables aleatorias normales, con distribuciones beta y gamma. Además, se podía disponer fácilmente de tablas para estas posteriores (en el apéndice o de fácil acceso con numerosos paquetes de software). Hay siempre un conjunto emergente de procedimientos de computadora en el que la posterior se determina con base en lo que introduzca el usuario acerca de la función de verosimilitud para los datos y la previa para el parámetro. Una vez obtenida la posterior por medio del software, se usa exactamente como lo describimos antes.

Los estimadores de Bayes se pueden evaluar por medio de criterios clásicos iterativos. Ya hemos visto que los estimadores de Bayes son sesgados. No obstante, suelen ser consistentes y, dependiendo de los criterios que se utilicen, pueden ser mejores que los estimadores iterativos correspondientes. En el Ejercicio 16.8, el lector determinó que la MSE del estimador de Bayes era a veces menor que la MSE de estimación de verosimilitud máxima (MLE). Además, la influencia de la selección de los valores de parámetro previos disminuye cuando el tamaño de la muestra aumenta.

En el Ejemplo 8.11 determinamos que el intervalo de confianza iterativo *realizado* para la media de una población normalmente distribuida era (2926.3, 2991.7). Con el uso de la perspectiva iterativa, la verdadera media poblacional es fija pero desconocida. En consecuencia, este intervalo *realizado* captura el verdadero valor de μ o no lo captura. Dijimos que este era un intervalo de confianza de 95% porque el procedimiento (fórmula) empleado para producirlo da intervalos que *capturan* la media fija alrededor de 95% del tiempo si muestras de tamaño 8 *se toman repetida e independientemente* y se usan para construir muchos intervalos. Si se toman y usan 100 muestras de tamaño 8 para producir intervalos de confianza realizados (diferentes), esperamos que aproximadamente 95 de ellos capturen el parámetro. No sabemos cuál de los 100 intervalos captura la media fija desconocida. La *misma información* se utilizó en el Ejemplo 16.5 para obtener (2946.89, 2967.57) como un intervalo *creíble* de 95% para μ, ahora visto como una variable aleatoria. Desde la perspectiva de Bayes, tiene sentido expresar que la probabilidad posterior es .95 de que la media (aleatoria) se incluya en este intervalo (fijo).

La bondad de las pruebas clásicas de hipótesis se mide por α y β, las probabilidades de errores tipo I y tipo II, respectivamente. Si pruebas con $\alpha = .05$ se realizan en repetidas ocasiones (usando muestras diferentes, seleccionadas de manera independiente), entonces cuando H_0 es verdadera, es rechazada 5% del tiempo. Si H_0 es realmente verdadera y 100 muestras del mismo tamaño se toman de manera independiente, esperamos rechazar la hipótesis nula (verdadera) unas cinco veces. No tiene sentido incluso tratar de calcular las probabilidades de las hipótesis. Desde la perspectiva de Bayes, el parámetro de interés es una variable aleatoria con distribución posterior deducida por el analista. El cálculo de las probabilidades posteriores para cada una de las hipótesis es completamente apropiado y es la base para la decisión de una prueba de Bayes.

¿Cuál es la mejor propuesta, la de Bayes o la iterativa? Es imposible dar una respuesta universal a esta pregunta. En algunas aplicaciones la propuesta de Bayes será mejor; en otras la propuesta iterativa es mejor.

Bibliografía y lecturas adicionales

Bayes, T. 1764. "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans. Roy. Soc.* 53, 370–418.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis,* 2d ed. New York: Springer-Verlag.

Box, G. E. P. 1980. "Sampling and Bayes' Inference in Scientific Modeling and Robustness," *J. of the Royal Statistical Society, Series A* 143, 383–430.

Box, G. E. P., and G. C. Tiao. 1992. *Bayesian Inference in Statistical Analysis.* New York: Wiley Classics.

Casella, G., and R. L. Berger. 2002. *Statistical Inference,* 2d ed. Pacific Grove, Calif.: Duxbury.

Hogg, R. V., J. W. McKean, and A. T. Craig. 2005. *Introduction to Mathematical Statistics,* 6th ed. Upper Saddle River, N.J.: Prentice Hall.

Kepner, J., and D. Wackerly. 2002. "Observations on the Effect of the Prior Distribution on the Predictive Distribution in Bayesian Inferences," *Journal of Applied Statistics* 29(5): 761–769.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics,* 3d ed. New York: McGraw-Hill.

Rice, J. A. 1995. *Mathematical Statistics and Data Analysis,* 2d ed. Belmont, Calif.: Duxbury.

APÉNDICE 1

Matrices y otros resultados matemáticos útiles

A1.1 Matrices y álgebra de matrices

La siguiente exposición representa un examen muy elemental y condensado de matrices y operaciones con matrices. Si usted busca una introducción más completa del tema, consulte los libros citados en la bibliografía indicada al final del Capítulo 11.

Definiremos una *matriz* como un conjunto rectangular (arreglo) de números reales e indicaremos matrices específicas simbólicamente con letras mayúsculas negritas. Los números en la matriz, *elementos*, aparecen en posiciones específicas de renglón-columna, todas las cuales se llenan. El número de renglones y columnas puede variar de una matriz a otra, de modo que en forma conveniente describimos el tamaño de una matriz al dar sus *dimensiones*, es decir, el número de renglones y columnas. Así, la matriz **A**

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 6 & 0 & -1 \\ 4 & 2 & 7 \end{bmatrix}$$

posee dimensiones 2 × 3 porque contiene dos renglones y tres columnas. Del mismo modo, para

$$\underset{4\times 1}{\mathbf{B}} = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 7 \end{bmatrix} \quad \text{y} \quad \underset{2\times 2}{\mathbf{C}} = \begin{bmatrix} 2 & 0 \\ -1 & 4 \end{bmatrix}$$

las dimensiones de **B** y **C** son 4 × 1 y 2 × 2, respectivamente. Observe que la dimensión de renglón siempre aparece primero y que las dimensiones se pueden escribir abajo del símbolo identificador de la matriz como se indica para las matrices **A, B** y **C.**

Al igual que en álgebra ordinaria, un elemento de una matriz puede estar indicado por un símbolo, *a, b,...*, y su posición renglón-columna puede estar identificado por medio de un doble subíndice. Así, a_{21} sería el elemento del segundo renglón, primera columna. Los renglones se numeran en orden de arriba abajo y las columnas de izquierda a derecha. En la matriz **A**, $a_{21} = 4$, $a_{13} = -1$, y así sucesivamente.

Los elementos de un renglón particular están identificados por el subíndice de su columna y, en consecuencia, están numerados de izquierda a derecha. El primer elemento de un renglón está a la izquierda. Análogamente, los elementos de una columna particular están identificados por el subíndice de su renglón y, por tanto, están identificados del elemento superior de la columna al inferior. Por ejemplo, el primer elemento de la columna 2 de la matriz **A** es 0, el segundo es 2. Los elementos primero, segundo y tercero del renglón 1 son 6, 0 y −1, respectivamente.

El término *álgebra de matrices* comprende, como su nombre lo indica, un álgebra que se refiere a matrices, en forma muy semejante a como el álgebra ordinaria se refiere a números o símbolos que representan números reales. Por tanto, expresaremos reglas para la suma y multiplicación de matrices, así definiremos otros elementos de un sistema algebraico. Al hacerlo, señalamos las similitudes o diferencias entre álgebra de matrices y álgebra ordinaria. Por último, usaremos nuestras operaciones con matrices para expresar y resolver una *ecuación matricial* muy sencilla. Ésta, como es de esperarse, será la solución que deseamos para las ecuaciones con mínimos cuadrados.

A1.2 Suma de matrices

Dos matrices, por ejemplo **A** y **B**, se pueden sumar *sólo* si son de las mismas dimensiones. La suma de dos matrices será una matriz obtenida al sumar elementos *correspondientes* de las matrices **A** y **B**, es decir, elementos en posiciones correspondientes. Siendo éste el caso, la suma resultante será una matriz de las mismas dimensiones que **A** y **B**.

EJEMPLO **A1.1** Encuentre la suma indicada de las matrices **A** y **B**:

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 2 & 1 & 4 \\ -1 & 6 & 0 \end{bmatrix} \qquad \underset{2\times 3}{\mathbf{B}} = \begin{bmatrix} 0 & -1 & 1 \\ 6 & -3 & 2 \end{bmatrix}$$

Solución

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}2 & 1 & 4\\ -1 & 6 & 0\end{bmatrix}+\begin{bmatrix}0 & -1 & 1\\ 6 & -3 & 2\end{bmatrix}$$

$$=\begin{bmatrix}(2+0) & (1-1) & (4+1)\\ (-1+6) & (6-3) & (0+2)\end{bmatrix}=\begin{bmatrix}2 & 0 & 5\\ 5 & 3 & 2\end{bmatrix}.$$

■

EJEMPLO **A1.2** Encuentre la suma de las matrices

$$\underset{3\times 3}{\mathbf{A}}=\begin{bmatrix}1 & 0 & 3\\ 1 & -1 & 4\\ 2 & -1 & 0\end{bmatrix} \quad \text{y} \quad \underset{3\times 3}{\mathbf{B}}=\begin{bmatrix}4 & 2 & -1\\ 1 & 0 & 6\\ 3 & 1 & 4\end{bmatrix}.$$

Solución ■

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}5 & 2 & 2\\ 2 & -1 & 10\\ 5 & 0 & 4\end{bmatrix}.$$

Observe que $(\mathbf{A}+\mathbf{B})=(\mathbf{B}+\mathbf{A})$, igual que en álgebra ordinaria y recuerde que nunca sumamos matrices de dimensiones diferentes.

A1.3 Multiplicación de una matriz por un número real

Deseamos una regla para multiplicar una matriz por un número real, por ejemplo, 3**A,** donde

$$\mathbf{A}=\begin{bmatrix}2 & 1\\ 4 & 6\\ -1 & 0\end{bmatrix}.$$

Ciertamente buscaríamos que 3**A** fuera $(\mathbf{A}+\mathbf{A}+\mathbf{A})$, para apegarnos a la regla de adición. Por tanto, 3**A** significaría que cada elemento de la matriz **A** debe multiplicarse por el multiplicador 3, y

$$3\mathbf{A}=\begin{bmatrix}3(2) & 3(1)\\ 3(4) & 3(6)\\ 3(-1) & 3(0)\end{bmatrix}=\begin{bmatrix}6 & 3\\ 12 & 18\\ -3 & 0\end{bmatrix}.$$

En general, dado un número real c y una matriz **A** con elementos a_{ij}, el producto $c\mathbf{A}$ será una matriz cuyos elementos son iguales a ca_{ij}.

A1.4 Multiplicación de matrices

La regla para multiplicación de matrices requiere "multiplicación de renglón-columna", que definiremos más adelante. El procedimiento puede parecer un poco complicado al novato pero no debe resultar demasiado difícil después de un poco de práctica. Ilustramos con un ejemplo.

Sean **A** y **B**

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix}.$$

Un elemento en el i-ésimo *renglón* y j-ésima *columna* del producto **AB** se obtiene al multiplicar el i-ésimo *renglón de* **A** por la j-ésima *columna de* **B.** De este modo, el elemento del primer renglón, primera columna de **AB** se obtiene multiplicando el primer renglón de **A** por la primera columna de **B**. De igual manera, el elemento del primer renglón, segunda columna sería el producto del primer renglón de **A** y la segunda columna de **B**. Observe que siempre usamos los renglones de **A** y las columnas de **B**, donde **A** es la matriz a la izquierda de **B** del producto **AB.**

La multiplicación renglón-columna es relativamente fácil. Obtenga los productos, elemento del primer renglón por elemento de primera columna, elemento del segundo renglón por elemento de segunda columna, tercero por tercera y luego sume. Recuerde que los elementos de renglón y columna se marcan de izquierda a derecha y de arriba abajo, respectivamente.

Si aplicamos estas reglas a nuestro ejemplo, obtenemos

$$\underset{2\times 2}{\mathbf{A}}\ \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 4 \\ 1 & 14 \end{bmatrix}.$$

El producto del primer renglón por la primera columna sería (2)(5) + (0)(−1) = 10, que está ubicado (y circulado) en el primer renglón, primera columna de **AB.** Del mismo modo, el elemento del primer renglón, segunda columna es igual al producto del primer renglón de **A** por la segunda columna de **B**, o (2)(2) + (0)(3) = 4. El producto del segundo renglón por la primera columna es (1)(5) + (4)(−1) = 1 y está situado en el segundo renglón, primera columna de **AB.** Por último, el producto del segundo renglón por la segunda columna es (1)(2) + (4)(3) = 14.

EJEMPLO **A1.3** Encuentre los productos **AB** y **BA**, donde

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix}.$$

Solución

$$\underset{3\times 2}{\mathbf{A}}\ \underset{2\times 3}{\mathbf{B}} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 10 & -2 & 0 \\ 2 & -1 & -3 \\ 8 & 0 & 8 \end{bmatrix}$$

y

$$\underset{2\times 3}{\mathbf{B}}\ \underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 1 \\ 4 & 10 \end{bmatrix}. \qquad \blacksquare$$

Observe que en álgebra matricial, a diferencia del álgebra ordinaria, **AB** no es igual a **BA**. Debido a que **A** contiene tres renglones y **B** tres columnas, podemos formar (3)(3) = 9 combinaciones renglón-columna por tanto nueve elementos para **AB.** En contraste, **B** contiene sólo dos renglones, **A** dos columnas y por tanto el producto **BA** tendrá sólo (2)(2) = 4 elementos, correspondientes a las cuatro combinaciones diferentes renglón-columna.

Además, observamos que la multiplicación renglón-columna se basa en la suposición de que los renglones de la matriz de la izquierda contienen el mismo número de elementos que las columnas de la matriz a la derecha, de modo que elementos correspondientes existirán para la multiplicación renglón-columna. ¿Qué hacemos cuando esta condición no se satisface? Convenimos en no multiplicar dos matrices, por ejemplo **AB,** donde los renglones de **A** y las columnas de **B** contienen un número desigual de elementos.

Un examen de las dimensiones de las matrices dirá si se pueden multiplicar así como dar las dimensiones del producto. Escribiendo las dimensiones bajo las dos matrices,

$$\underset{m\times p}{\mathbf{A}}\ \underset{p\times q}{\mathbf{B}} = \underset{m\times q}{\mathbf{AB}}$$

observamos que los dos números interiores, que dan el número de elementos en un renglón de **A** y columna de **B**, respectivamente, deben ser iguales. Los dos números exteriores, que indican el número de renglones de **A** y columnas de **B**, dan las dimensiones de la matriz producto. Se puede verificar la operación de esta regla para el Ejemplo A1.3.

EJEMPLO **A1.4** Obtenga el producto **AB:**

$$\underset{1\times 3}{\mathbf{A}}\ \underset{3\times 2}{\mathbf{B}} = [2 \quad 1 \quad 0]\begin{bmatrix} 2 & 0 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} = [4 \quad 3]$$

Observe que el producto **AB** es (1 × 2) y que **BA** no está definida por las respectivas dimensiones de **A** y **B**.

EJEMPLO **A1.5** Encuentre el producto **AB**, donde

$$\mathbf{A} = [1 \quad 2 \quad 3 \quad 4] \qquad \text{y} \qquad \mathbf{B} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Solución

$$\underset{1\times 4}{\mathbf{A}}\ \underset{4\times 1}{\mathbf{B}} = [1 \quad 2 \quad 3 \quad 4]\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = [30].$$

Observe que este ejemplo produce un método diferente para escribir una suma de cuadrados. ■

A1.5 Elementos identidad

Los elementos identidad para adición y multiplicación en álgebra ordinaria son 0 y 1, respectivamente. En adición, 0 más cualquier otro elemento, por ejemplo a, es igual a a; esto es,

$$0 + 2 = 2, \qquad 0 + (-9) = -9.$$

Análogamente, la multiplicación del elemento identidad 1 por cualquier otro elemento, a por ejemplo, es igual a a; esto es,

$$(1)(5) = 5, \qquad (1)(-4) = -4.$$

En álgebra de matrices, se dice que dos matrices son iguales cuando todos los elementos correspondientes son iguales. Con esto en mente definiremos las matrices de identidad de un modo semejante al empleado en álgebra ordinaria. Por tanto, si $\mathbf{A}$ es cualquier matriz, una matriz $\mathbf{B}$ será una matriz identidad para adición si

$$\mathbf{A} + \mathbf{B} = \mathbf{A} \qquad \text{y} \qquad \mathbf{B} + \mathbf{A} = \mathbf{A}.$$

Con facilidad se puede ver que la matriz identidad para adición es aquella en la que todo elemento es igual a cero. Esta matriz es de interés pero no tiene importancia práctica en nuestro trabajo.

Del mismo modo, si $\mathbf{A}$ es cualquier matriz, la matriz identidad para multiplicación es una matriz $\mathbf{I}$ que satisface la relación

$$\mathbf{AI} = \mathbf{A} \qquad \text{e} \qquad \mathbf{IA} = \mathbf{A}.$$

Esta matriz, llamada *matriz identidad*, es la *matriz cuadrada*

$$\underset{n\times n}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Esto es, todos los elementos de la *diagonal principal* de la matriz, que corren de arriba a la izquierda a abajo a la derecha, son iguales a 1; todos los otros elementos son iguales a cero. Observe que la matriz identidad está siempre indicada por el símbolo $\mathbf{I}$.

A diferencia del álgebra ordinaria, que contiene sólo un elemento identidad para multiplicación, el álgebra de matrices debe contener un número infinitamente grande de matrices identidad. Así, debemos tener matrices con dimensiones 1×1, 2×2, 3×3, 4×4, etcétera, para dar una identidad de las dimensiones correctas para permitir multiplicación. Todo será de esta forma.

Que la matriz $\mathbf{I}$ satisfaga la relación

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}$$

se puede demostrar en un ejemplo.

EJEMPLO **A1.6** Sea

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 6 & 3 \end{bmatrix}.$$

Demuestre que $\mathbf{IA} = \mathbf{A}$ y que $\mathbf{AI} = \mathbf{A}$.

Solución

$$\underset{2\times 2}{\mathbf{I}}\ \underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ -1 & 6 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 6 & 3 \end{bmatrix} = \mathbf{A}$$

y

$$\underset{2\times 3}{\mathbf{A}}\ \underset{3\times 3}{\mathbf{I}} = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 6 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 6 & 3 \end{bmatrix} = \mathbf{A}.$$

■

A1.6 La inversa de una matriz

Para que el álgebra de matrices sea útil, debemos ser capaces de construir y resolver ecuaciones de matrices para una matriz de incógnitas en forma semejante a la empleada en álgebra ordinaria. Esto, a su vez, requiere un método de efectuar una división.

Por ejemplo, resolveríamos la ecuación sencilla en álgebra ordinaria,

$$2x = 6$$

al dividir entre 2 ambos lados de la ecuación y obtener $x = 3$. Otra forma de ver esta operación es para definir el recíproco de cada elemento, en un sistema algebraico y considerar la división como multiplicación por el recíproco de un elemento. Podríamos resolver la ecuación $2x = 6$ al multiplicar ambos lados de la ecuación por el recíproco de 2. En vista de que todo elemento del sistema de números reales posee un recíproco, con excepción de 0, la operación de multiplicación elimina la necesidad de la división.

El recíproco de un número c en álgebra ordinaria es un número b que satisface la relación

$$cb = 1$$

esto es, el producto de un número por su recíproco debe ser igual al elemento identidad para multiplicación. Por ejemplo, el recíproco de 2 es $1/2$ y $(2)(1/2) = 1$.

Un recíproco en álgebra de matrices recibe el nombre de *inverso* de una matriz y se define como sigue:

DEFINICIÓN **A1.1**

Sea $\mathbf{A}_{n\times n}$ una matriz cuadrada. Si se puede hallar una matriz $\mathbf{A}^{-1}$ tal que

$$\mathbf{A}\mathbf{A}^{-1} = 1 \qquad \text{y} \qquad \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

entonces $\mathbf{A}^{-1}$ recibe el nombre de *inversa* de **A.**

Observe que el requisito para una inversa en álgebra de matrices es el mismo que en álgebra ordinaria, es decir, el producto de **A** por su inversa debe ser igual a la matriz identidad para multiplicación. Además, la inversa no está definida para matrices no cuadradas y, por tanto, numerosas matrices en álgebra de matrices no tienen inversas (recuerde que 0 era el único elemento del sistema de números reales sin inversa). Por último, expresamos sin prueba que muchas matrices cuadradas no tienen inversas; las que sí tienen se identificarán en la Sección A1.9 y se dará un método para hallar la inversa de una matriz.

A1.7 La transpuesta de una matriz

Hemos explicado la relación que hay entre una matriz y su inversa. Una segunda relación matricial útil define la *transpuesta* de una matriz.

DEFINICIÓN **A1.2**

Sea $A_{p\times q}$ una matriz de dimensiones $p \times q$. Entonces $\mathbf{A}'$, llamada la *transpuesta* de **A**, se define como la matriz obtenida al intercambiar renglones y columnas correspondientes de **A**; esto es, primera con primera, segunda con segunda, etcétera.

Por ejemplo, sea

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 4 & 3 \end{bmatrix}.$$

Entonces

$$\underset{2\times 3}{\mathbf{A}'} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & 3 \end{bmatrix}.$$

Observe que los renglones primero y segundo de $\mathbf{A}'$ son idénticos a la primera y segunda columnas, respectivamente, de **A**.

Como segundo ejemplo, sea

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

Entonces $\mathbf{Y}' = [y_1\ y_2\ y_3]$. Como punto de interés, observamos que $\mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{3} y_i^2$.

Por último, si

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & 3 \\ 1 & 6 & 9 \end{bmatrix}$$

entonces

$$\mathbf{A}' = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 2 & 6 \\ 4 & 3 & 9 \end{bmatrix}.$$

A1.8 Una expresión matricial para un sistema de ecuaciones lineales simultáneas

A continuación presentamos una de las aplicaciones sencillas e importantes del álgebra de matrices. Sea

$$2v_1 + v_2 = 5$$

$$v_1 - v_2 = 1$$

un par de ecuaciones lineales simultáneas con dos variables, v_1 y v_2. Entonces definimos tres matrices:

$$\underset{2\times2}{\mathbf{A}} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \qquad \underset{2\times1}{\mathbf{V}} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad \underset{2\times1}{\mathbf{G}} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Observe que **A** es la matriz de los coeficientes de las incógnitas, cuando las ecuaciones se escriben cada una de ellas con las variables apareciendo en el mismo orden, leyendo de izquierda a derecha y con las constantes en el lado derecho del signo de igualdad. La matriz **V** proporciona las incógnitas en una columna y en el mismo orden en que aparecen en las ecuaciones. Por último, la matriz **G** contiene las constantes en una columna exactamente como aparecen en el conjunto de ecuaciones.

El sistema simultáneo de dos ecuaciones lineales puede ahora escribirse en notación de matrices como

$$\mathbf{AV} = \mathbf{G}$$

este enunciado puede ser verificado fácilmente al multiplicar **A** y **V** y luego comparar la respuesta con **G**.

$$\mathbf{AV} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 + v_2 \\ v_1 - v_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix} = G.$$

Observe que elementos correspondientes en **AV** y **G** son iguales, es decir, $2v_1 + v_2 = 5$ y $v_1 - v_2 = 1$. Por tanto, **AV** = **G**.

El método para escribir un par de ecuaciones lineales con dos incógnitas como una ecuación de matrices se puede extender fácilmente a un sistema de r ecuaciones con r incógnitas. Por ejemplo, si las ecuaciones son

$$\begin{aligned}
a_{11}v_1 + a_{12}v_2 + a_{13}v_3 + \cdots + a_{1r}v_r &= g_1 \\
a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + \cdots + a_{2r}v_r &= g_2 \\
a_{31}v_1 + a_{32}v_2 + a_{33}v_3 + \cdots + a_{3r}v_r &= g_3 \\
\vdots \quad\quad \vdots \quad\quad \vdots \quad\quad\quad \vdots \;\; &= \;\vdots \\
a_{r1}v_1 + a_{r2}v_2 + a_{r3}v_3 + \cdots + a_{rr}v_r &= g_r
\end{aligned}$$

defina

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1r} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2r} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3r} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & a_{r3} & \cdots & a_{rr} \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_r \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ g_r \end{bmatrix}.$$

Observe que, una vez más, **A** es una matriz cuadrada de coeficientes variables, mientras que **V** y **G** son matrices de columna que contienen las variables y constantes, respectivamente. Entonces **AV** = **G**.

Cualquiera que sea el tamaño del sistema de ecuaciones, si tenemos n ecuaciones lineales con n incógnitas, el sistema se puede escribir como la ecuación matricial simple **AV** = **G**.

Usted observará que la matriz **V** contiene todas las incógnitas, mientras que **A** y **G** son matrices constantes.

Nuestro objetivo, por supuesto, es despejar la matriz de incógnitas, $\mathbf{V}$, donde la ecuación $\mathbf{AV} = \mathbf{G}$ es semejante a la ecuación

$$2v = 6$$

en álgebra ordinaria. Si esto es verdadero, no es de sorprenderse hallar que los métodos de soluciones son los mismos. En álgebra ordinaria, ambos lados de la ecuación se multiplican por el recíproco de 2; en álgebra de matrices, ambos lados de la ecuación se multiplican por $\mathbf{A}^{-1}$. Entonces

$$\mathbf{A}^{-1}(\mathbf{AV}) = \mathbf{A}^{-1}\mathbf{G}$$

o bien,

$$\mathbf{A}^{-1}\mathbf{AV} = \mathbf{A}^{-1}\mathbf{G}.$$

Pero $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ y además $\mathbf{IV} = \mathbf{V}$. Por tanto, $\mathbf{V} = \mathbf{A}^{-1}\mathbf{G}$. En otras palabras, las soluciones al sistema de ecuaciones lineales simultáneas se pueden obtener al hallar $\mathbf{A}^{-1}$ y luego obtener el producto $\mathbf{A}^{-1}\mathbf{G}$. Los valores de solución de $v_1, v_2, v_3, \ldots, v_r$ aparecerán en sucesión en la matriz de columna $\mathbf{V} = \mathbf{A}^{-1}\mathbf{G}$.

A1.9 Inversión de una matriz

Hemos indicado en la Sección A1.8 que la clave para las soluciones de un sistema de ecuaciones lineales simultáneas por el método de álgebra de matrices se apoya en la adquisición de la inversa de la matriz **A.** Existen numerosos métodos para invertir matrices. El método que presentamos no es el mejor desde un punto de vista computacional, pero funciona muy bien para las matrices asociadas con la mayor parte de diseños experimentales y es uno de los más fáciles de presentar al estudiante novato. Depende de un teorema en álgebra de matrices y el uso de *operaciones de renglón.*

Antes de definir *operaciones de renglón* en matrices, debemos expresar lo que se quiere decir por adición de dos renglones de una matriz y la multiplicación de un renglón por una constante. Ilustraremos con la matriz $\mathbf{A}$ para el sistema de dos ecuaciones lineales simultáneas,

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}.$$

Dos renglones de una matriz se pueden sumar si se suman sus elementos correspondientes. Así, si los dos renglones de la matriz $\mathbf{A}$ se suman, se obtiene un nuevo renglón con elementos $[(2 + 1)(1 - 1)] = [3\ 0]$. La multiplicación de un renglón por una constante significa que cada elemento del renglón se multiplica por la constante. Dos veces el primer renglón de la matriz $\mathbf{A}$ generaría el renglón $[4\ 2]$. Con estas ideas en mente, definiremos tres formas de operar en un renglón de una matriz:

1. Un renglón puede ser multiplicado por una constante.
2. Un renglón puede ser multiplicado por una constante y sumado o restado de otro renglón (que está identificado como aquel en el que se realiza la operación).
3. Dos renglones pueden intercambiarse.

Dada la matriz $\mathbf{A}$, es muy fácil ver que podríamos efectuar una serie de operaciones de renglón que darían la misma nueva matriz $\mathbf{B}$. En este sentido expresamos sin demostración un sorprendente e interesante teorema de álgebra de matrices; es decir, existe alguna matriz **C** tal que

$$\mathbf{CA} = \mathbf{B}$$

En otras palabras, una serie de operaciones de renglón en una matriz **A** es equivalente a multiplicar **A** por una matriz **C**. Usaremos este principio para invertir una matriz.

Coloque la matriz **A**, que ha de invertirse, junto a una matriz identidad de las mismas dimensiones:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

A continuación realice las mismas operaciones de renglón en **A** *e* **I** *en forma tal que* **A** *cambie a una matriz identidad.* Al hacerlo, debemos haber multiplicado **A** por una matriz **C** de modo que $\mathbf{CA} = \mathbf{I}$. Por tanto, ¡**C** debe ser la inversa de **A**! El problema, desde luego, es hallar la matriz desconocida **C** y, por fortuna, esto resulta ser de muy poca dificultad. Debido a que realizamos las mismas operaciones de renglón en **A** e **I**, la matriz identidad debe haber cambiado a $\mathbf{CI} = \mathbf{C} = \mathbf{A}^{-1}$.

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\downarrow \text{(mismas operaciones de renglón)} \downarrow$$

$$\mathbf{CA} = \mathbf{I} \qquad \mathbf{CI} = \mathbf{C} = \mathbf{A}^{-1}$$

Ilustraremos con el siguiente ejemplo.

EJEMPLO **A1.7** Invierta la matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}.$$

Solución

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Paso 1. Trabaje en el renglón 1 al multiplicar el renglón 1 por 1/2. (Nota: Es útil para el estudiante novato identificar el renglón en el que esté operando porque todos los otros renglones seguirán sin cambio, aun cuando pueden usarse en la operación. Pondremos un asterisco en el renglón sobre el cual la operación se realice.)

$$\begin{matrix} * \\ \end{matrix}\begin{bmatrix} 1 & 1/2 \\ 1 & -1 \end{bmatrix} \quad \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Paso 2. Trabaje en el renglón 2 al restar el renglón 1 del renglón 2.

$$\begin{matrix} \\ * \end{matrix}\begin{bmatrix} 1 & 1/2 \\ 0 & -3/2 \end{bmatrix} \quad \begin{bmatrix} 1/2 & 0 \\ -1/2 & 1 \end{bmatrix}.$$

(Observe que el renglón 2 se usa simplemente para operar en el renglón 1 y por tanto permanece sin cambio.)

Paso 3. Multiplique el renglón 2 por (−2/3).

$$\begin{matrix} \\ * \end{matrix}\begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1/2 & 0 \\ 1/3 & -2/3 \end{bmatrix}.$$

Paso 4. Trabaje en el renglón 1 al multiplicar el renglón 2 por 1/2 y restar del renglón 1.

$$* \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix}.$$

(Observe que el renglón 2 se usa simplemente para operar en el renglón 1 y por tanto permanece sin cambio.) En consecuencia, el inverso de **A** debe ser

$$\mathbf{A}^{-1} = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix}.$$

Existe una verificación fácil de los cálculos para el procedimiento de inversión porque $\mathbf{A}^{-1}\,\mathbf{A}$ debe ser igual a la matriz identidad **I**. Entonces

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ ■

EJEMPLO **A1.8** Invierta la matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 2 \\ 1 & 0 & 0 \end{bmatrix}$$

y compruebe los resultados.

Solución

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Paso 1. Multiplique el renglón 1 por $1/2$.

$$\begin{matrix} * \\ \\ \\ \end{matrix} \begin{bmatrix} 1 & 0 & 1/2 \\ 1 & -1 & 2 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Paso 2. Trabaje en el renglón 2 al restar el renglón 1 del renglón 2.

$$\begin{matrix} \\ * \\ \\ \end{matrix} \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & -1 & 3/2 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1/2 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Paso 3. Trabaje en el renglón 3 al restar el renglón 1 del renglón 3.

$$\begin{matrix} \\ \\ * \end{matrix} \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & -1 & 3/2 \\ 0 & 0 & -1/2 \end{bmatrix} \qquad \begin{bmatrix} 1/2 & 0 & 0 \\ -1/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix}.$$

Paso 4. Trabaje en el renglón 2 al multiplicar el renglón 3 por 3 y sumar al renglón 2.

$$\begin{matrix} \\ * \\ \\ \end{matrix} \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & -1 & 0 \\ 0 & 0 & -1/2 \end{bmatrix} \qquad \begin{bmatrix} 1/2 & 0 & 0 \\ -2 & 1 & 3 \\ -1/2 & 0 & 1 \end{bmatrix}.$$

Paso 5. Multiplique el renglón 2 por (−1).

$$*\begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{bmatrix} \quad \begin{bmatrix} 1/2 & 0 & 0 \\ 2 & -1 & -3 \\ -1/2 & 0 & 1 \end{bmatrix}.$$

Paso 6. Trabaje en el renglón 1 al sumar el renglón 3 al renglón 1.

$$*\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 \\ 2 & -1 & -3 \\ -1/2 & 0 & 1 \end{bmatrix}.$$

Paso 7. Multiplique el renglón 3 por (−2).

$$*\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 \\ 2 & -1 & -3 \\ 1 & 0 & -2 \end{bmatrix} = \mathbf{A}^{-1}.$$

Las siete operaciones de renglón han cambiado la matriz **A** a la matriz identidad y, excluyendo errores de cálculo, han cambiado la identidad a $\mathbf{A}^{-1}$.

Al comprobar, tenemos

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 2 & -1 & -3 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 2 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

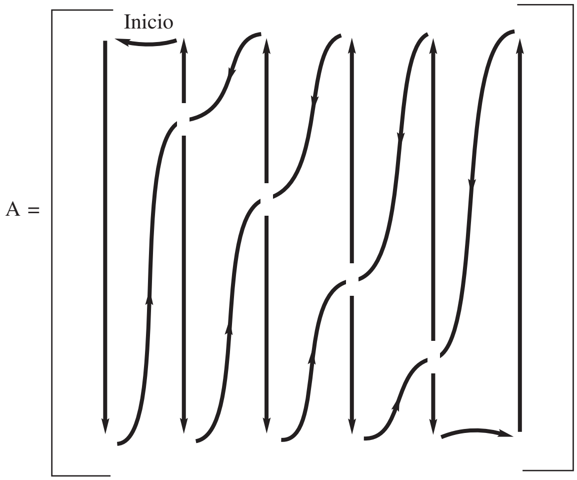
Vemos que $\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}$ y por tanto que los cálculos son correctos. ■

Obsérvese que la sucesión de operaciones de renglón requeridas para convertir **A** en **I** no es única. Una persona podría obtener la inversa si usa cinco operaciones de renglón, mientras que otra podría necesitar diez, pero el resultado final será el mismo. No obstante, para mayor eficiencia es deseable emplear un sistema.

Observe que el proceso de inversión utiliza operaciones de renglón para cambiar elementos fuera de la diagonal de la matriz **A** a números 0 y los elementos de la diagonal principal a números 1. Un procedimiento sistemático es el siguiente. Cambie a 1 el elemento de arriba a la izquierda y luego realice operaciones de renglón para cambiar a 0 todos los otros elementos de la *primera* columna. A continuación pase al elemento diagonal del segundo renglón, segunda columna, cámbielo a un 1, y cambie a 0 todos los elementos de la *segunda* columna *debajo de* la diagonal principal. Este proceso se repite, al moverse hacia la parte inferior de la diagonal principal de arriba a la izquierda a abajo a la derecha, hasta que todos los elementos debajo de la diagonal principal hayan cambiado a 0. Para eliminar elementos diferentes de cero arriba de la diagonal principal, trabaje en todos los elementos de la última columna, cambiando cada uno de ellos a 0; entonces pase al siguiente a la última columna y repita el proceso. Continúe este procedimiento hasta que llegue al primer elemento de la primera columna, que fue el punto de partida. Este procedimiento está indicado en forma diagramática en la Figura A1.1.

La inversión de matrices es un proceso tedioso, en el mejor de los casos, y requiere tanto trabajo como las soluciones de un sistema de ecuaciones simultáneas por eliminación o sustitución. Al lector le agradará saber que no esperamos que desarrolle facilidad para inversión de matrices. Por fortuna, casi todas las matrices asociadas con experimentos diseñados siguen patrones y se invierten fácilmente.

FIGURA **A1.1** Procedimiento para inversión de matrices



Será benéfico que el lector invierta unas cuantas matrices de 2×2 y de 3×3. Las matrices que carecen de un modelo, en particular las matrices grandes, se invierten de manera más eficiente y económica en computadora. (Se han perfeccionado programas para inversión de matrices para casi todas las computadoras.)

Destacamos el hecho de que obtener las soluciones para las ecuaciones de mínimos cuadrados (Capítulo 11) por inversión de matrices tiene ventajas distintivas que pueden o no ser evidentes. En especial está el hecho de que el procedimiento de inversión es sistemático y, por tanto, es particularmente apropiado para cálculo en computadora. No obstante, la principal ventaja es que el procedimiento de inversión de manera automática producirá las varianzas de los estimadores de todos los parámetros del modelo lineal.

Antes de salir del tema de inversión de matrices, preguntamos cómo se puede identificar una matriz que tenga una inversa. Una consulta a una explicación de ecuaciones lineales de álgebra ordinaria debe dar la respuesta.

Evidentemente, una solución única para un sistema de ecuaciones lineales simultáneas no puede obtenerse a menos que las ecuaciones sean independientes. Entonces, si una de las ecuaciones es una combinación lineal de las otras, las ecuaciones son dependientes. Las matrices de coeficientes asociadas con sistemas dependientes de ecuaciones lineales no poseen una inversa.

A1.10 Resolución de un sistema de ecuaciones lineales simultáneas

Finalmente hemos obtenido todos los ingredientes necesarios para resolver un sistema de ecuaciones lineales simultáneas,

$$
\begin{aligned}
2v_1 + v_2 &= 5 \\
v_1 - v_2 &= 1
\end{aligned}
$$

Si recordamos que las soluciones matriciales al sistema de ecuaciones $\mathbf{AV} = \mathbf{G}$ es $\mathbf{V} = \mathbf{A}^{-1}\mathbf{G}$, obtenemos

$$\mathbf{V} = \mathbf{A}^{-1}\mathbf{G} = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Por tanto, la solución es

$$\mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

esto es, $v_1 = 2$ y $v_2 = 1$, lo cual se puede verificar por sustitución de estos valores en las ecuaciones lineales originales.

EJEMPLO **A1.9** Resuelva el sistema de ecuaciones lineales simultáneas

$$\begin{aligned} 2v_1 + v_3 &= 4 \\ v_1 - v_2 + 2v_3 &= 2 \\ v_1 &= 1. \end{aligned}$$

Solución La matriz coeficiente para estas ecuaciones,

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1 \\ 1 & -1 & 2 \\ 1 & 0 & 0 \end{bmatrix}$$

apareció en el Ejemplo A1.8. En ese ejemplo, encontramos que

$$\mathbf{A}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 2 & -1 & -3 \\ 1 & 0 & -2 \end{bmatrix}.$$

Resolviendo, tenemos

$$\mathbf{V} = \mathbf{A}^{-1}\mathbf{G} = \begin{bmatrix} 0 & 0 & 1 \\ 2 & -1 & -3 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$

Entonces, $v_1 = 1$, $v_2 = 3$ y $v_3 = 2$ da las soluciones al conjunto de tres ecuaciones lineales simultáneas.

■

A1.11 Otros resultados matemáticos útiles

El propósito de esta sección es hacer que usted cuente con una referencia conveniente para algunos de los resultados matemáticos clave que se usan con frecuencia en el cuerpo del texto.

La expansión binomial de $(x + y)^n$ Sean x e y cualesquier números reales, entonces

$$
\begin{aligned}
(x+y)^n &= \binom{n}{0} x^n y^0 + \binom{n}{1} x^{n-1} y^1 + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n} x^0 y^n \\
&= \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i .
\end{aligned}
$$

La suma de una serie geométrica Sea r un número real tal que $|r| < 1$, y m sea cualquier entero $m \geq 1$

$$
\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}, \quad \sum_{i=1}^{\infty} r^i = \frac{r}{1-r}, \quad \sum_{i=0}^{m} r^i = \frac{1-r^{m+1}}{1-r}.
$$

La expansión de la serie (de Taylor) de e^x Sea x cualquier número real, entonces

$$
e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}.
$$

A continuación tenemos algunas fórmulas útiles para sumas particulares. Las pruebas (omitidas) se establecen con más facilidad con el uso de inducción matemática.

$$
\sum_{i=1}^{n} i = \frac{n(n+1)}{2}
$$

Función gamma Sea $t > 0$, entonces $\Gamma(t)$ está definida por la integral siguiente:

$$
\Gamma(t) = \int_0^{\infty} y^{t-1} e^{-y} dy.
$$

Usando la técnica de integración por partes, se deduce que para cualquier $t > 0$

$$
\Gamma(t+1) = t\Gamma(t)
$$

y si $t = n$, donde n es un entero,

$$
\Gamma(n) = (n-1)!.
$$

Además,

$$
\Gamma(1/2) = \sqrt{\pi}.
$$

Si $\alpha, \beta > 0$, la ***función beta***, $B(\alpha, \beta)$, está definida por la siguiente integral,

$$
B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy
$$

y está relacionada a la función gamma como sigue:

$$
B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.
$$

APÉNDICE 2

Distribuciones, medias, varianzas y funciones generadoras de momento de probabilidad común

Tabla 1 Distribuciones discretas

Distribución	Función de probabilidad	Media	Varianza	Función generadora de momento
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$; $y = 0, 1, \ldots, n$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geométrica	$p(y) = p(1-p)^{y-1}$; $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hipergeométrica	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$; $y = 0, 1, \ldots, n$ si $n \leq r$, $y = 0, 1, \ldots, r$ si $n > r$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	No existe en forma cerrada
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$; $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$
Binomial negativa	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$; $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

Tabla 2 Distribuciones continuas

Distribución	Función de probabilidad	Media	Varianza	Función generadora de momento
Uniforme	$f(y) = \frac{1}{\theta_2 - \theta_1} : \theta_1 \leq y \leq \theta_2$	$\frac{\theta_1 + \theta_2}{2}$	$\frac{(\theta_2 - \theta_1)^2}{12}$	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$ $-\infty < y < +\infty$	μ	σ^2	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Exponencial	$f(y) = \frac{1}{\beta} e^{-y/\beta} : \quad \beta > 0$ $0 < y < \infty$	β	β^2	$(1 - \beta t)^{-1}$
Gamma	$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{\alpha-1} e^{-y/\beta};$ $0 < y < \infty$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Ji-cuadrada	$f(y) = \frac{(y)^{(v/2)-1} e^{-y/2}}{2^{v/2}\Gamma(v/2)};$ $y > 0$	v	$2v$	$(1 - 2t)^{-v/2}$
Beta	$f(y) = \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] y^{\alpha-1}(1-y)^{\beta-1};$ $0 < y < 1$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$	no existe en forma cerrada

APÉNDICE 3

Tablas

Tabla 1 Probabilidades binomiales

Los valores tabulados son $P(Y \leq a) = \sum_{y=0}^{a} p(y)$. (Los cálculos están redondeados al tercer lugar decimal.)
(a) $n = 5$

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*a*
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4

(b) $n = 10$

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*a*
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000	0
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000	1
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000	2
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000	3
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000	4
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000	5
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000	6
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000	7
8	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086	.004	8
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096	9

Tabla 1 (*continúa*)

(c) $n = 15$

a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.860	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.990	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000	.000	1
2	1.000	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000	.000	2
3	1.000	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000	.000	3
4	1.000	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000	.000	4
5	1.000	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000	.000	5
6	1.000	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000	.000	6
7	1.000	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000	.000	7
8	1.000	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000	.000	8
9	1.000	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000	.000	9
10	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001	.000	10
11	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005	.000	11
12	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036	.000	12
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171	.010	13
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537	.140	14

(d) $n = 20$

a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.818	.358	.122	.012	.001	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.983	.736	.392	.069	.008	.001	.000	.000	.000	.000	.000	.000	.000	1
2	.999	.925	.677	.206	.035	.004	.000	.000	.000	.000	.000	.000	.000	2
3	1.000	.984	.867	.411	.107	.016	.001	.000	.000	.000	.000	.000	.000	3
4	1.000	.997	.957	.630	.238	.051	.006	.000	.000	.000	.000	.000	.000	4
5	1.000	1.000	.989	.804	.416	.126	.021	.002	.000	.000	.000	.000	.000	5
6	1.000	1.000	.998	.913	.608	.250	.058	.006	.000	.000	.000	.000	.000	6
7	1.000	1.000	1.000	.968	.772	.416	.132	.021	.001	.000	.000	.000	.000	7
8	1.000	1.000	1.000	.990	.887	.596	.252	.057	.005	.000	.000	.000	.000	8
9	1.000	1.000	1.000	.997	.952	.755	.412	.128	.017	.001	.000	.000	.000	9
10	1.000	1.000	1.000	.999	.983	.872	.588	.245	.048	.003	.000	.000	.000	10
11	1.000	1.000	1.000	1.000	.995	.943	.748	.404	.113	.010	.000	.000	.000	11
12	1.000	1.000	1.000	1.000	.999	.979	.868	.584	.228	.032	.000	.000	.000	12
13	1.000	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.087	.002	.000	.000	13
14	1.000	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.196	.011	.000	.000	14
15	1.000	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.370	.043	.003	.000	15
16	1.000	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.589	.133	.016	.000	16
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.794	.323	.075	.001	17
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.931	.608	.264	.017	18
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.988	.878	.642	.182	19

Tabla 1 (***continúa***)

(e) $n = 25$

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.778	.277	.072	.004	.000	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.974	.642	.271	.027	.002	.000	.000	.000	.000	.000	.000	.000	.000	1
2	.998	.873	.537	.098	.009	.000	.000	.000	.000	.000	.000	.000	.000	2
3	1.000	.966	.764	.234	.033	.002	.000	.000	.000	.000	.000	.000	.000	3
4	1.000	.993	.902	.421	.090	.009	.000	.000	.000	.000	.000	.000	.000	4
5	1.000	.999	.967	.617	.193	.029	.002	.000	.000	.000	.000	.000	.000	5
6	1.000	1.000	.991	.780	.341	.074	.007	.000	.000	.000	.000	.000	.000	6
7	1.000	1.000	.998	.891	.512	.154	.022	.001	.000	.000	.000	.000	.000	7
8	1.000	1.000	1.000	.953	.677	.274	.054	.004	.000	.000	.000	.000	.000	8
9	1.000	1.000	1.000	.983	.811	.425	.115	.013	.000	.000	.000	.000	.000	9
10	1.000	1.000	1.000	.994	.902	.586	.212	.034	.002	.000	.000	.000	.000	10
11	1.000	1.000	1.000	.998	.956	.732	.345	.078	.006	.000	.000	.000	.000	11
12	1.000	1.000	1.000	1.000	.983	.846	.500	.154	.017	.000	.000	.000	.000	12
13	1.000	1.000	1.000	1.000	.994	.922	.655	.268	.044	.002	.000	.000	.000	13
14	1.000	1.000	1.000	1.000	.998	.966	.788	.414	.098	.006	.000	.000	.000	14
15	1.000	1.000	1.000	1.000	1.000	.987	.885	.575	.189	.017	.000	.000	.000	15
16	1.000	1.000	1.000	1.000	1.000	.996	.946	.726	.323	.047	.000	.000	.000	16
17	1.000	1.000	1.000	1.000	1.000	.999	.978	.846	.488	.109	.002	.000	.000	17
18	1.000	1.000	1.000	1.000	1.000	1.000	.993	.926	.659	.220	.009	.000	.000	18
19	1.000	1.000	1.000	1.000	1.000	1.000	.998	.971	.807	.383	.033	.001	.000	19
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.910	.579	.098	.007	.000	20
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.967	.766	.236	.034	.000	21
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.902	.463	.127	.002	22
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.973	.729	.358	.026	23
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.928	.723	.222	24

Tabla 2 Tabla de e^{-x}

x	e^{-x}	x	e^{-x}	x	e^{-x}	x	e^{-x}
0.00	1.000000	2.60	.074274	5.10	.006097	7.60	.000501
0.10	.904837	2.70	.067206	5.20	.005517	7.70	.000453
0.20	.818731	2.80	.060810	5.30	.004992	7.80	.000410
0.30	.740818	2.90	.055023	5.40	.004517	7.90	.000371
0.40	.670320	3.00	.049787	5.50	.004087	8.00	.000336
0.50	.606531	3.10	.045049	5.60	.003698	8.10	.000304
0.60	.548812	3.20	.040762	5.70	.003346	8.20	.000275
0.70	.496585	3.30	.036883	5.80	.003028	8.30	.000249
0.80	.449329	3.40	.033373	5.90	.002739	8.40	.000225
0.90	.406570	3.50	.030197	6.00	.002479	8.50	.000204
1.00	.367879	3.60	.027324	6.10	.002243	8.60	.000184
1.10	.332871	3.70	.024724	6.20	.002029	8.70	.000167
1.20	.301194	3.80	.022371	6.30	.001836	8.80	.000151
1.30	.272532	3.90	.020242	6.40	.001661	8.90	.000136
1.40	.246597	4.00	.018316	6.50	.001503	9.00	.000123
1.50	.223130	4.10	.016573	6.60	.001360	9.10	.000112
1.60	.201897	4.20	.014996	6.70	.001231	9.20	.000101
1.70	.182684	4.30	.013569	6.80	.001114	9.30	.000091
1.80	.165299	4.40	.012277	6.90	.001008	9.40	.000083
1.90	.149569	4.50	.011109	7.00	.000912	9.50	.000075
2.00	.135335	4.60	.010052	7.10	.000825	9.60	.000068
2.10	.122456	4.70	.009095	7.20	.000747	9.70	.000061
2.20	.110803	4.80	.008230	7.30	.000676	9.80	.000056
2.30	.100259	4.90	.007447	7.40	.000611	9.90	.000050
2.40	.090718	5.00	.006738	7.50	.000553	10.00	.000045
2.50	.082085						

Tabla 3 Probabilidades de Poisson

$$P(Y \le a) = \sum_{y=0}^{a} e^{-\lambda} \frac{\lambda^y}{y!}$$

λ \ a	0	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	0.999	1.000						
0.20	0.819	0.982	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.994	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.55	0.577	0.894	0.982	0.988	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.65	0.522	0.861	0.972	0.996	0.999	1.000				
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.75	0.472	0.827	0.959	0.993	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.85	0.427	0.791	0.945	0.989	0.998	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
0.95	0.387	0.754	0.929	0.981	0.997	1.000				
1.00	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.998	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.833	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.995	0.999	1.000	

Tabla 3 (*continúa*)

λ \ a	0	1	2	3	4	5	6	7	8	9
2.2	0.111	0.355	0.623	0.819	0.928	0.975	0.993	0.998	1.000	
2.4	0.091	0.308	0.570	0.779	0.904	0.964	0.988	0.997	0.999	1.000
2.6	0.074	0.267	0.518	0.736	0.877	0.951	0.983	0.995	0.999	1.000
2.8	0.061	0.231	0.469	0.692	0.848	0.935	0.976	0.992	0.998	0.999
3.0	0.050	0.199	0.423	0.647	0.815	0.916	0.966	0.988	0.996	0.999
3.2	0.041	0.171	0.380	0.603	0.781	0.895	0.955	0.983	0.994	0.998
3.4	0.033	0.147	0.340	0.558	0.744	0.871	0.942	0.977	0.992	0.997
3.6	0.027	0.126	0.303	0.515	0.706	0.844	0.927	0.969	0.988	0.996
3.8	0.022	0.107	0.269	0.473	0.668	0.816	0.909	0.960	0.984	0.994
4.0	0.018	0.092	0.238	0.433	0.629	0.785	0.889	0.949	0.979	0.992
4.2	0.015	0.078	0.210	0.395	0.590	0.753	0.867	0.936	0.972	0.989
4.4	0.012	0.066	0.185	0.359	0.551	0.720	0.844	0.921	0.964	0.985
4.6	0.010	0.056	0.163	0.326	0.513	0.686	0.818	0.905	0.955	0.980
4.8	0.008	0.048	0.143	0.294	0.476	0.651	0.791	0.887	0.944	0.975
5.0	0.007	0.040	0.125	0.265	0.440	0.616	0.762	0.867	0.932	0.968
5.2	0.006	0.034	0.109	0.238	0.406	0.581	0.732	0.845	0.918	0.960
5.4	0.005	0.029	0.095	0.213	0.373	0.546	0.702	0.822	0.903	0.951
5.6	0.004	0.024	0.082	0.191	0.342	0.512	0.670	0.797	0.886	0.941
5.8	0.003	0.021	0.072	0.170	0.313	0.478	0.638	0.771	0.867	0.929
6.0	0.002	0.017	0.062	0.151	0.285	0.446	0.606	0.744	0.847	0.916

λ	10	11	12	13	14	15	16
2.8	1.000						
3.0	1.000						
3.2	1.000						
3.4	0.999	1.000					
3.6	0.999	1.000					
3.8	0.998	0.999	1.000				
4.0	0.997	0.999	1.000				
4.2	0.996	0.999	1.000				
4.4	0.994	0.998	0.999	1.000			
4.6	0.992	0.997	0.999	1.000			
4.8	0.990	0.996	0.999	1.000			
5.0	0.986	0.995	0.998	0.999	1.000		
5.2	0.982	0.993	0.997	0.999	1.000		
5.4	0.977	0.990	0.996	0.999	1.000		
5.6	0.972	0.988	0.995	0.998	0.999	1.000	
5.8	0.965	0.984	0.993	0.997	0.999	1.000	
6.0	0.957	0.980	0.991	0.996	0.999	0.999	1.000

Tabla 3 (*continúa*)

λ \ a	0	1	2	3	4	5	6	7	8	9
6.2	0.002	0.015	0.054	0.134	0.259	0.414	0.574	0.716	0.826	0.902
6.4	0.002	0.012	0.046	0.119	0.235	0.384	0.542	0.687	0.803	0.886
6.6	0.001	0.010	0.040	0.105	0.213	0.355	0.511	0.658	0.780	0.869
6.8	0.001	0.009	0.034	0.093	0.192	0.327	0.480	0.628	0.755	0.850
7.0	0.001	0.007	0.030	0.082	0.173	0.301	0.450	0.599	0.729	0.830
7.2	0.001	0.006	0.025	0.072	0.156	0.276	0.420	0.569	0.703	0.810
7.4	0.001	0.005	0.022	0.063	0.140	0.253	0.392	0.539	0.676	0.788
7.6	0.001	0.004	0.019	0.055	0.125	0.231	0.365	0.510	0.648	0.765
7.8	0.000	0.004	0.016	0.048	0.112	0.210	0.338	0.481	0.620	0.741
8.0	0.000	0.003	0.014	0.042	0.100	0.191	0.313	0.453	0.593	0.717
8.5	0.000	0.002	0.009	0.030	0.074	0.150	0.256	0.386	0.523	0.653
9.0	0.000	0.001	0.006	0.021	0.055	0.116	0.207	0.324	0.456	0.587
9.5	0.000	0.001	0.004	0.015	0.040	0.089	0.165	0.269	0.392	0.522
10.0	0.000	0.000	0.003	0.010	0.029	0.067	0.130	0.220	0.333	0.458
	10	11	12	13	14	15	16	17	18	19
6.2	0.949	0.975	0.989	0.995	0.998	0.999	1.000			
6.4	0.939	0.969	0.986	0.994	0.997	0.999	1.000			
6.6	0.927	0.963	0.982	0.992	0.997	0.999	0.999	1.000		
6.8	0.915	0.955	0.978	0.990	0.996	0.998	0.999	1.000		
7.0	0.901	0.947	0.973	0.987	0.994	0.998	0.999	1.000		
7.2	0.887	0.937	0.967	0.984	0.993	0.997	0.999	0.999	1.000	
7.4	0.871	0.926	0.961	0.980	0.991	0.996	0.998	0.999	1.000	
7.6	0.854	0.915	0.954	0.976	0.989	0.995	0.998	0.999	1.000	
7.8	0.835	0.902	0.945	0.971	0.986	0.993	0.997	0.999	1.000	
8.0	0.816	0.888	0.936	0.966	0.983	0.992	0.996	0.998	0.999	1.000
8.5	0.763	0.849	0.909	0.949	0.973	0.986	0.993	0.997	0.999	0.999
9.0	0.706	0.803	0.876	0.926	0.959	0.978	0.989	0.995	0.998	0.999
9.5	0.645	0.752	0.836	0.898	0.940	0.967	0.982	0.991	0.996	0.998
10.0	0.583	0.697	0.792	0.864	0.917	0.951	0.973	0.986	0.993	0.997
	20	21	22							
8.5	1.000									
9.0	1.000									
9.5	0.999	1.000								
10.0	0.998	0.999	1.000							

Tabla 3 (*continúa*)

λ \ a	0	1	2	3	4	5	6	7	8	9
10.5	0.000	0.000	0.002	0.007	0.021	0.050	0.102	0.179	0.279	0.397
11.0	0.000	0.000	0.001	0.005	0.015	0.038	0.079	0.143	0.232	0.341
11.5	0.000	0.000	0.001	0.003	0.011	0.028	0.060	0.114	0.191	0.289
12.0	0.000	0.000	0.001	0.002	0.008	0.020	0.046	0.090	0.155	0.242
12.5	0.000	0.000	0.000	0.002	0.005	0.015	0.035	0.070	0.125	0.201
13.0	0.000	0.000	0.000	0.001	0.004	0.011	0.026	0.054	0.100	0.166
13.5	0.000	0.000	0.000	0.001	0.003	0.008	0.019	0.041	0.079	0.135
14.0	0.000	0.000	0.000	0.000	0.002	0.006	0.014	0.032	0.062	0.109
14.5	0.000	0.000	0.000	0.000	0.001	0.004	0.010	0.024	0.048	0.088
15.0	0.000	0.000	0.000	0.000	0.001	0.003	0.008	0.018	0.037	0.070
	10	11	12	13	14	15	16	17	18	19
10.5	0.521	0.639	0.742	0.825	0.888	0.932	0.960	0.978	0.988	0.994
11.0	0.460	0.579	0.689	0.781	0.854	0.907	0.944	0.968	0.982	0.991
11.5	0.402	0.520	0.633	0.733	0.815	0.878	0.924	0.954	0.974	0.986
12.0	0.347	0.462	0.576	0.682	0.772	0.844	0.899	0.937	0.963	0.979
12.5	0.297	0.406	0.519	0.628	0.725	0.806	0.869	0.916	0.948	0.969
13.0	0.252	0.353	0.463	0.573	0.675	0.764	0.835	0.890	0.930	0.957
13.5	0.211	0.304	0.409	0.518	0.623	0.718	0.798	0.861	0.908	0.942
14.0	0.176	0.260	0.358	0.464	0.570	0.669	0.756	0.827	0.883	0.923
14.5	0.145	0.220	0.311	0.413	0.518	0.619	0.711	0.790	0.853	0.901
15.0	0.118	0.185	0.268	0.363	0.466	0.568	0.664	0.749	0.819	0.875
	20	21	22	23	24	25	26	27	28	29
10.5	0.997	0.999	0.999	1.000						
11.0	0.995	0.998	0.999	1.000						
11.5	0.992	0.996	0.998	0.999	1.000					
12.0	0.988	0.994	0.997	0.999	0.999	1.000				
12.5	0.983	0.991	0.995	0.998	0.999	0.999	1.000			
13.0	0.975	0.986	0.992	0.996	0.998	0.999	1.000			
13.5	0.965	0.980	0.989	0.994	0.997	0.998	0.999	1.000		
14.0	0.952	0.971	0.983	0.991	0.995	0.997	0.999	0.999	1.000	
14.5	0.936	0.960	0.976	0.986	0.992	0.996	0.998	0.999	0.999	1.000
15.0	0.917	0.947	0.967	0.981	0.989	0.994	0.997	0.998	0.999	1.000

Tabla 3 (*continúa*)

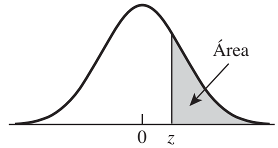
λ \ a	4	5	6	7	8	9	10	11	12	13
16	0.000	0.001	0.004	0.010	0.022	0.043	0.077	0.127	0.193	0.275
17	0.000	0.001	0.002	0.005	0.013	0.026	0.049	0.085	0.135	0.201
18	0.000	0.000	0.001	0.003	0.007	0.015	0.030	0.055	0.092	0.143
19	0.000	0.000	0.001	0.002	0.004	0.009	0.018	0.035	0.061	0.098
20	0.000	0.000	0.000	0.001	0.002	0.005	0.011	0.021	0.039	0.066
21	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.013	0.025	0.043
22	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.008	0.015	0.028
23	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.009	0.017
24	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.011
25	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.006

	14	15	16	17	18	19	20	21	22	23
16	0.368	0.467	0.566	0.659	0.742	0.812	0.868	0.911	0.942	0.963
17	0.281	0.371	0.468	0.564	0.655	0.736	0.805	0.861	0.905	0.937
18	0.208	0.287	0.375	0.469	0.562	0.651	0.731	0.799	0.855	0.899
19	0.150	0.215	0.292	0.378	0.469	0.561	0.647	0.725	0.793	0.849
20	0.105	0.157	0.221	0.297	0.381	0.470	0.559	0.644	0.721	0.787
21	0.072	0.111	0.163	0.227	0.302	0.384	0.471	0.558	0.640	0.716
22	0.048	0.077	0.117	0.169	0.232	0.306	0.387	0.472	0.556	0.637
23	0.031	0.052	0.082	0.123	0.175	0.238	0.310	0.389	0.472	0.555
24	0.020	0.034	0.056	0.087	0.128	0.180	0.243	0.314	0.392	0.473
25	0.012	0.022	0.038	0.060	0.092	0.134	0.185	0.247	0.318	0.394

	24	25	26	27	28	29	30	31	32	33
16	0.978	0.987	0.993	0.996	0.998	0.999	0.999	1.000		
17	0.959	0.975	0.985	0.991	0.995	0.997	0.999	0.999	1.000	
18	0.932	0.955	0.972	0.983	0.990	0.994	0.997	0.998	0.999	1.000
19	0.893	0.927	0.951	0.969	0.980	0.988	0.993	0.996	0.998	0.999
20	0.843	0.888	0.922	0.948	0.966	0.978	0.987	0.992	0.995	0.997
21	0.782	0.838	0.883	0.917	0.944	0.963	0.976	0.985	0.991	0.994
22	0.712	0.777	0.832	0.877	0.913	0.940	0.959	0.973	0.983	0.989
23	0.635	0.708	0.772	0.827	0.873	0.908	0.936	0.956	0.971	0.981
24	0.554	0.632	0.704	0.768	0.823	0.868	0.904	0.932	0.953	0.969
25	0.473	0.553	0.629	0.700	0.763	0.818	0.863	0.900	0.929	0.950

	34	35	36	37	38	39	40	41	42	43
19	0.999	1.000								
20	0.999	0.999	1.000							
21	0.997	0.998	0.999	0.999	1.000					
22	0.994	0.996	0.998	0.999	0.999	1.000				
23	0.988	0.993	0.996	0.997	0.999	0.999	1.000			
24	0.979	0.987	0.992	0.995	0.997	0.998	0.999	0.999	1.000	
25	0.966	0.978	0.985	0.991	0.991	0.997	0.998	0.999	0.999	1.000

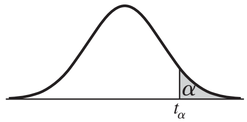
Tabla 4 Áreas de curva normal
Probabilidad normal estándar en cola derecha
(para valores negativos de z, las áreas se encuentran por simetría)



	Segundo lugar decimal de z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

De R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

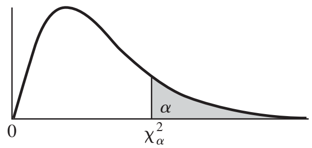
Tabla 5 Puntos porcentuales de las distribuciones *t*



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	gl
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15
1.337	1.746	2.120	2.583	2.921	16
1.333	1.740	2.110	2.567	2.898	17
1.330	1.734	2.101	2.552	2.878	18
1.328	1.729	2.093	2.539	2.861	19
1.325	1.725	2.086	2.528	2.845	20
1.323	1.721	2.080	2.518	2.831	21
1.321	1.717	2.074	2.508	2.819	22
1.319	1.714	2.069	2.500	2.807	23
1.318	1.711	2.064	2.492	2.797	24
1.316	1.708	2.060	2.485	2.787	25
1.315	1.706	2.056	2.479	2.779	26
1.314	1.703	2.052	2.473	2.771	27
1.313	1.701	2.048	2.467	2.763	28
1.311	1.699	2.045	2.462	2.756	29
1.282	1.645	1.960	2.326	2.576	inf.

De "Table of Percentage Points of the *t*-Distribution". Calculado por Maxine Merrington, *Biometrika*, Vol. 32 (1941), p. 300.

Tabla 6 Puntos porcentuales de las distribuciones χ^2



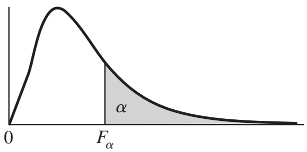
gl	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

Tabla 6 (***continúa***)

$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$	gl
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8454	31.9999	34.2672	16
24.7690	27.5871	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
51.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1539	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100

De "Table of the Percentage Points of the χ^2-Distribution". Calculado por Maxine Merrington, *Biometrika*, Vol. 32 (1941), pp.188-189, por Catherine M. Thompson.

Tabla 7 Puntos porcentuales de las distribuciones *F*



GL del denominador	GL del numerador									
	α	1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
	.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
	.010	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
	.005	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51

Tabla 7 (*continúa*)

F_α

GL del numerador											
10	12	15	20	24	30	40	60	120	∞	α	gl
60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33	.100	1
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3	.050	
968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018	.025	
6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	.010	
24224	24426	24630	24836	24940	25044	25148	25253	25359	25465	.005	
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	.100	2
19.40	19.41	19.43	19.45	19.45	19.45	19.47	19.48	19.49	19.50	.050	
39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50	.025	
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	.010	
199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	.005	
5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13	.100	3
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	.050	
14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	.025	
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	.010	
43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83	.005	
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	.100	4
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	.050	
8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26	.025	
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	.010	
20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32	.005	
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10	.100	5
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	.050	
6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	.025	
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	.010	
13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14	.005	
2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	.100	6
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	.050	
5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	.025	
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	.010	
10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88	.005	
2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	.100	7
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	.050	
4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	.025	
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	.010	
8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08	.005	

Tabla 7 (*continúa*)

F_α

GL del denominador	GL del numerador									
	α	1	2	3	4	5	6	7	8	9
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72

Tabla 7 (*continúa*)

F_α

GL del numerador											
10	12	15	20	24	30	40	60	120	∞	α	gl
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	.100	8
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	.050	
4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	.025	
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	.010	
7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95	.005	
2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	.100	9
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	.050	
3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	.025	
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	.010	
6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19	.005	
2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06	.100	10
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	.050	
3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	.025	
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	.010	
5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64	.005	
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97	.100	11
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	.050	
3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	.025	
4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	.010	
5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23	.005	
2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90	.100	12
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	.050	
3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	.025	
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	.010	
5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90	.005	
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85	.100	13
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	.050	
3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	.025	
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	.010	
4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65	.005	
2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80	.100	14
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	.050	
3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	.025	
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	.010	
4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44	.005	

Tabla 7 **(*continúa*)**

F_α

GL del denominador	GL del numerador									
	α	1	2	3	4	5	6	7	8	9
15	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
	.005	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.005	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88

Tabla 7 (*continúa*)

F_α

GL del numerador											
10	12	15	20	24	30	40	60	120	∞	α	gl
2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76	.100	15
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	.050	
3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40	.025	
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	.010	
4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26	.005	
2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72	.100	16
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	.050	
2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	.025	
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	.010	
4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11	.005	
2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69	.100	17
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	.050	
2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	.025	
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	.010	
4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98	.005	
1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66	.100	18
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	.050	
2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	.025	
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	.010	
4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87	.005	
1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63	.100	19
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	.050	
2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	.025	
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	.010	
3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78	.005	
1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61	.100	20
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	.050	
2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	.025	
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	.010	
3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69	.005	
1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59	.100	21
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	.050	
2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04	.025	
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	.010	
3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61	.005	

Tabla 7 (***continúa***)

F_α

GL del denominador	GL del numerador									
	α	1	2	3	4	5	6	7	8	9
22	.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.005	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.005	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.005	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	.025	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.005	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
	.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	.005	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52

Tabla 7 (*continúa*)

F_α

GL del numerador											
10	12	15	20	24	30	40	60	120	∞	α	gl
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57	.100	22
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	.050	
2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00	.025	
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	.010	
3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55	.005	
1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55	.100	23
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	.050	
2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97	.025	
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	.010	
3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48	.005	
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53	.100	24
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	.050	
2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	.025	
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	.010	
3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43	.005	
1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	.100	25
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	.050	
2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	.025	
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	.010	
3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38	.005	
1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	.100	26
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	.050	
2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	.025	
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	.010	
3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33	.005	
1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	.100	27
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	.050	
2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	.025	
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	.010	
3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.29	.005	
1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	.100	28
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	.050	
2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	.025	
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	.010	
3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.25	.005	

Tabla 7 (***continúa***)

F_α

GL del denominador	GL del numerador									
	α	1	2	3	4	5	6	7	8	9
29	.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.025	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	.005	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.005	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	.100	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	.050	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
	.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	.100	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

De "Tables of percentage points of the inverted beta (F) distribution". *Biometrika*, Vol. 33 (1943) por M. Merrington y C. M. Thompson y de la Tabla 18 de *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, 1954, editado por E. S. Pearson y H. O. Hartley.

Tabla 7 **(*continúa*)**

F_α

GL del numerador											
10	12	15	20	24	30	40	60	120	∞	α	gl
1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	.100	29
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	.050	
2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	.025	
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	.010	
3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.21	.005	
1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	.100	30
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	.050	
2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	.025	
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	.010	
3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18	.005	
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	.100	40
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	.050	
2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	.025	
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	.010	
3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93	.005	
1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	.100	60
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	.050	
2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	.025	
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	.010	
2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69	.005	
1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19	.100	120
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25	.050	
2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	.025	
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	.010	
2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43	.005	
1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00	.100	∞
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	.050	
2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	.025	
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	.010	
2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00	.005	

Tabla 8 Función de distribución de *U*

$P(U \leq U_0)$; U_0 es el argumento; $n_1 \leq n_2$; $3 \leq n_2 \leq 10$.

$n_2 = 3$

	n_1		
U_0	1	2	3
0	.25	.10	.05
1	.50	.20	.10
2		.40	.20
3		.60	.35
4			.50

$n_2 = 4$

	n_1			
U_0	1	2	3	4
0	.2000	.0667	.0286	.0143
1	.4000	.1333	.0571	.0286
2	.6000	.2667	.1143	.0571
3		.4000	.2000	.1000
4		.6000	.3143	.1714
5			.4286	.2429
6			.5714	.3429
7				.4429
8				.5571

Tabla 8 (*continúa*)

$n_2 = 5$

	n_1				
U_0	1	2	3	4	5
0	.1667	.0476	.0179	.0079	.0040
1	.3333	.0952	.0357	.0159	.0079
2	.5000	.1905	.0714	.0317	.0159
3		.2857	.1250	.0556	.0278
4		.4286	.1964	.0952	.0476
5		.5714	.2857	.1429	.0754
6			.3929	.2063	.1111
7			.5000	.2778	.1548
8				.3651	.2103
9				.4524	.2738
10				.5476	.3452
11					.4206
12					.5000

$n_2 = 6$

	n_1					
U_0	1	2	3	4	5	6
0	.1429	.0357	.0119	.0048	.0022	.0011
1	.2857	.0714	.0238	.0095	.0043	.0022
2	.4286	.1429	.0476	.0190	.0087	.0043
3	.5714	.2143	.0833	.0333	.0152	.0076
4		.3214	.1310	.0571	.0260	.0130
5		.4286	.1905	.0857	.0411	.0206
6		.5714	.2738	.1286	.0628	.0325
7			.3571	.1762	.0887	.0465
8			.4524	.2381	.1234	.0660
9			.5476	.3048	.1645	.0898
10				.3810	.2143	.1201
11				.4571	.2684	.1548
12				.5429	.3312	.1970
13					.3961	.2424
14					.4654	.2944
15					.5346	.3496
16						.4091
17						.4686
18						.5314

Tabla 8 (*continúa*)

$n_2 = 7$

	n_1						
U_0	1	2	3	4	5	6	7
0	.1250	.0278	.0083	.0030	.0013	.0006	.0003
1	.2500	.0556	.0167	.0061	.0025	.0012	.0006
2	.3750	.1111	.0333	.0121	.0051	.0023	.0012
3	.5000	.1667	.0583	.0212	.0088	.0041	.0020
4		.2500	.0917	.0364	.0152	.0070	.0035
5		.3333	.1333	.0545	.0240	.0111	.0055
6		.4444	.1917	.0818	.0366	.0175	.0087
7		.5556	.2583	.1152	.0530	.0256	.0131
8			.3333	.1576	.0745	.0367	.0189
9			.4167	.2061	.1010	.0507	.0265
10			.5000	.2636	.1338	.0688	.0364
11				.3242	.1717	.0903	.0487
12				.3939	.2159	.1171	.0641
13				.4636	.2652	.1474	.0825
14				.5364	.3194	.1830	.1043
15					.3775	.2226	.1297
16					.4381	.2669	.1588
17					.5000	.3141	.1914
18						.3654	.2279
19						.4178	.2675
20						.4726	.3100
21						.5274	.3552
22							.4024
23							.4508
24							.5000

Tabla 8 (*continúa*)

$n_2 = 8$

	n_1							
U_0	1	2	3	4	5	6	7	8
0	.1111	.0222	.0061	.0020	.0008	.0003	.0002	.0001
1	.2222	.0444	.0121	.0040	.0016	.0007	.0003	.0002
2	.3333	.0889	.0242	.0081	.0031	.0013	.0006	.0003
3	.4444	.1333	.0424	.0141	.0054	.0023	.0011	.0005
4	.5556	.2000	.0667	.0242	.0093	.0040	.0019	.0009
5		.2667	.0970	.0364	.0148	.0063	.0030	.0015
6		.3556	.1394	.0545	.0225	.0100	.0047	.0023
7		.4444	.1879	.0768	.0326	.0147	.0070	.0035
8		.5556	.2485	.1071	.0466	.0213	.0103	.0052
9			.3152	.1414	.0637	.0296	.0145	.0074
10			.3879	.1838	.0855	.0406	.0200	.0103
11			.4606	.2303	.1111	.0539	.0270	.0141
12			.5394	.2848	.1422	.0709	.0361	.0190
13				.3414	.1772	.0906	.0469	.0249
14				.4040	.2176	.1142	.0603	.0325
15				.4667	.2618	.1412	.0760	.0415
16				.5333	.3108	.1725	.0946	.0524
17					.3621	.2068	.1159	.0652
18					.4165	.2454	.1405	.0803
19					.4716	.2864	.1678	.0974
20					.5284	.3310	.1984	.1172
21						.3773	.2317	.1393
22						.4259	.2679	.1641
23						.4749	.3063	.1911
24						.5251	.3472	.2209
25							.3894	.2527
26							.4333	.2869
27							.4775	.3227
28							.5225	.3605
29								.3992
30								.4392
31								.4796
32								.5204

Tabla 8 (*continúa*)

$n_2 = 9$

	n_1								
U_0	1	2	3	4	5	6	7	8	9
0	.1000	.0182	.0045	.0014	.0005	.0002	.0001	.0000	.0000
1	.2000	.0364	.0091	.0028	.0010	.0004	.0002	.0001	.0000
2	.3000	.0727	.0182	.0056	.0020	.0008	.0003	.0002	.0001
3	.4000	.1091	.0318	.0098	.0035	.0014	.0006	.0003	.0001
4	.5000	.1636	.0500	.0168	.0060	.0024	.0010	.0005	.0002
5		.2182	.0727	.0252	.0095	.0038	.0017	.0008	.0004
6		.2909	.1045	.0378	.0145	.0060	.0026	.0012	.0006
7		.3636	.1409	.0531	.0210	.0088	.0039	.0019	.0009
8		.4545	.1864	.0741	.0300	.0128	.0058	.0028	.0014
9		.5455	.2409	.0993	.0415	.0180	.0082	.0039	.0020
10			.3000	.1301	.0559	.0248	.0115	.0056	.0028
11			.3636	.1650	.0734	.0332	.0156	.0076	.0039
12			.4318	.2070	.0949	.0440	.0209	.0103	.0053
13			.5000	.2517	.1199	.0567	.0274	.0137	.0071
14				.3021	.1489	.0723	.0356	.0180	.0094
15				.3552	.1818	.0905	.0454	.0232	.0122
16				.4126	.2188	.1119	.0571	.0296	.0157
17				.4699	.2592	.1361	.0708	.0372	.0200
18				.5301	.3032	.1638	.0869	.0464	.0252
19					.3497	.1942	.1052	.0570	.0313
20					.3986	.2280	.1261	.0694	.0385
21					.4491	.2643	.1496	.0836	.0470
22					.5000	.3035	.1755	.0998	.0567
23						.3445	.2039	.1179	.0680
24						.3878	.2349	.1383	.0807
25						.4320	.2680	.1606	.0951
26						.4773	.3032	.1852	.1112
27						.5227	.3403	.2117	.1290
28							.3788	.2404	.1487
29							.4185	.2707	.1701
30							.4591	.3029	.1933
31							.5000	.3365	.2181
32								.3715	.2447
33								.4074	.2729
34								.4442	.3024
35								.4813	.3332
36								.5187	.3652
37									.3981
38									.4317
39									.4657
40									.5000

Tabla 8 (*continúa*)

$n_2 = 10$

	n_1									
U_0	1	2	3	4	5	6	7	8	9	10
0	.0909	.0152	.0035	.0010	.0003	.0001	.0001	.0000	.0000	.0000
1	.1818	.0303	.0070	.0020	.0007	.0002	.0001	.0000	.0000	.0000
2	.2727	.0606	.0140	.0040	.0013	.0005	.0002	.0001	.0000	.0000
3	.3636	.0909	.0245	.0070	.0023	.0009	.0004	.0002	.0001	.0000
4	.4545	.1364	.0385	.0120	.0040	.0015	.0006	.0003	.0001	.0001
5	.5455	.1818	.0559	.0180	.0063	.0024	.0010	.0004	.0002	.0001
6		.2424	.0804	.0270	.0097	.0037	.0015	.0007	.0003	.0002
7		.3030	.1084	.0380	.0140	.0055	.0023	.0010	.0005	.0002
8		.3788	.1434	.0529	.0200	.0080	.0034	.0015	.0007	.0004
9		.4545	.1853	.0709	.0276	.0112	.0048	.0022	.0011	.0005
10		.5455	.2343	.0939	.0376	.0156	.0068	.0031	.0015	.0008
11			.2867	.1199	.0496	.0210	.0093	.0043	.0021	.0010
12			.3462	.1518	.0646	.0280	.0125	.0058	.0028	.0014
13			.4056	.1868	.0823	.0363	.0165	.0078	.0038	.0019
14			.4685	.2268	.1032	.0467	.0215	.0103	.0051	.0026
15			.5315	.2697	.1272	.0589	.0277	.0133	.0066	.0034
16				.3177	.1548	.0736	.0351	.0171	.0086	.0045
17				.3666	.1855	.0903	.0439	.0217	.0110	.0057
18				.4196	.2198	.1099	.0544	.0273	.0140	.0073
19				.4725	.2567	.1317	.0665	.0338	.0175	.0093
20				.5275	.2970	.1566	.0806	.0416	.0217	.0116
21					.3393	.1838	.0966	.0506	.0267	.0144
22					.3839	.2139	.1148	.0610	.0326	.0177
23					.4296	.2461	.1349	.0729	.0394	.0216
24					.4765	.2811	.1574	.0864	.0474	.0262
25					.5235	.3177	.1819	.1015	.0564	.0315
26						.3564	.2087	.1185	.0667	.0376
27						.3962	.2374	.1371	.0782	.0446
28						.4374	.2681	.1577	.0912	.0526
29						.4789	.3004	.1800	.1055	.0615
30						.5211	.3345	.2041	.1214	.0716
31							.3698	.2299	.1388	.0827
32							.4063	.2574	.1577	.0952
33							.4434	.2863	.1781	.1088
34							.4811	.3167	.2001	.1237
35							.5189	.3482	.2235	.1399
36								.3809	.2483	.1575
37								.4143	.2745	.1763
38								.4484	.3019	.1965
39								.4827	.3304	.2179

Tabla 8 (*continúa*)

$n_2 = 10$

	n_1									
U_0	1	2	3	4	5	6	7	8	9	10
40								.5173	.3598	.2406
41									.3901	.2644
42									.4211	.2894
43									.4524	.3153
44									.4841	.3421
45									.5159	.3697
46										.3980
47										.4267
48										.4559
49										.4853
50										.5147

Calculado por M. Pagano, Department of Statistics, University of Florida.

Tabla 9 Valores críticos de *T* en los pares acoplados de Wilcoxon: Prueba de rangos con signo, *n* = 5(1)50

Un lado	Dos lados	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$P = .05$	$P = .10$	1	2	4	6	8	11
$P = .025$	$P = .05$		1	2	4	6	8
$P = .01$	$P = .02$			0	2	3	5
$P = .005$	$P = .01$				0	2	3
Un lado	**Dos lados**	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$P = .05$	$P = .10$	14	17	21	26	30	36
$P = .025$	$P = .05$	11	14	17	21	25	30
$P = .01$	$P = .02$	7	10	13	16	20	24
$P = .005$	$P = .01$	5	7	10	13	16	19
Un lado	**Dos lados**	$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$P = .05$	$P = .10$	41	47	54	60	68	75
$P = .025$	$P = .05$	35	40	46	52	59	66
$P = .01$	$P = .02$	28	33	38	43	49	56
$P = .005$	$P = .01$	23	28	32	37	43	49
Un lado	**Dos lados**	$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$P = .05$	$P = .10$	83	92	101	110	120	130
$P = .025$	$P = .05$	73	81	90	98	107	117
$P = .01$	$P = .02$	62	69	77	85	93	102
$P = .005$	$P = .01$	55	68	68	76	84	92

Tabla 9 (*continúa*)

Un lado	Dos lados	$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$
$P = .05$	$P = .10$	141	152	163	175	188	201
$P = .025$	$P = .05$	127	137	148	159	171	183
$P = .01$	$P = .02$	111	120	130	141	151	162
$P = .005$	$P = .01$	100	109	118	128	138	149
Un lado	Dos lados	$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	
$P = .05$	$P = .10$	214	228	242	256	271	
$P = .025$	$P = .05$	195	208	222	235	250	
$P = .01$	$P = .02$	174	186	198	211	224	
$P = .005$	$P = .01$	160	171	183	195	208	
Un lado	Dos lados	$n = 40$	$n = 41$	$n = 42$	$n = 43$	$n = 44$	$n = 45$
$P = .05$	$P = .10$	287	303	319	336	353	371
$P = .025$	$P = .05$	264	279	295	311	327	344
$P = .01$	$P = .02$	238	252	267	281	297	313
$P = .005$	$P = .01$	221	234	248	262	277	292
Un lado	Dos lados	$n = 46$	$n = 47$	$n = 48$	$n = 49$	$n = 50$	
$P = .05$	$P = .10$	389	408	427	446	466	
$P = .025$	$P = .05$	361	379	397	415	434	
$P = .01$	$P = .02$	329	345	362	380	398	
$P = .005$	$P = .01$	307	323	339	356	373	

De “Some Rapid Approximate Statistical Procedures” (1964), 28, F. Wilcoxon y R. A. Wilcox.

Tabla 10 Distribución del número total de corridas R en muestras de tamaño (n_1, n_2), $P(R \leq a)$

	a								
(n_1, n_2)	2	3	4	5	6	7	8	9	10
(2, 3)	.200	.500	.900	1.000					
(2, 4)	.133	.400	.800	1.000					
(2, 5)	.095	.333	.714	1.000					
(2, 6)	.071	.286	.643	1.000					
(2, 7)	.056	.250	.583	1.000					
(2, 8)	.044	.222	.533	1.000					
(2, 9)	.036	.200	.491	1.000					
(2, 10)	.030	.182	.455	1.000					
(3, 3)	.100	.300	.700	.900	1.000				
(3, 4)	.057	.200	.543	.800	.971	1.000			
(3, 5)	.036	.143	.429	.714	.929	1.000			
(3, 6)	.024	.107	.345	.643	.881	1.000			
(3, 7)	.017	.083	.283	.583	.833	1.000			
(3, 8)	.012	.067	.236	.533	.788	1.000			
(3, 9)	.009	.055	.200	.491	.745	1.000			
(3, 10)	.007	.045	.171	.455	.706	1.000			
(4, 4)	.029	.114	.371	.629	.886	.971	1.000		
(4, 5)	.016	.071	.262	.500	.786	.929	.992	1.000	
(4, 6)	.010	.048	.190	.405	.690	.881	.976	1.000	
(4, 7)	.006	.033	.142	.333	.606	.833	.954	1.000	
(4, 8)	.004	.024	.109	.279	.533	.788	.929	1.000	
(4, 9)	.003	.018	.085	.236	.471	.745	.902	1.000	
(4, 10)	.002	.014	.068	.203	.419	.706	.874	1.000	
(5, 5)	.008	.040	.167	.357	.643	.833	.960	.992	1.000
(5, 6)	.004	.024	.110	.262	.522	.738	.911	.976	.998
(5, 7)	.003	.015	.076	.197	.424	.652	.854	.955	.992
(5, 8)	.002	.010	.054	.152	.347	.576	.793	.929	.984
(5, 9)	.001	.007	.039	.119	.287	.510	.734	.902	.972
(5, 10)	.001	.005	.029	.095	.239	.455	.678	.874	.958
(6, 6)	.002	.013	.067	.175	.392	.608	.825	.933	.987
(6, 7)	.001	.008	.043	.121	.296	.500	.733	.879	.966
(6, 8)	.001	.005	.028	.086	.226	.413	.646	.821	.937
(6, 9)	.000	.003	.019	.063	.175	.343	.566	.762	.902
(6, 10)	.000	.002	.013	.047	.137	.288	.497	.706	.864
(7, 7)	.001	.004	.025	.078	.209	.383	.617	.791	.922
(7, 8)	.000	.002	.015	.051	.149	.296	.514	.704	.867
(7, 9)	.000	.001	.010	.035	.108	.231	.427	.622	.806
(7, 10)	.000	.001	.006	.024	.080	.182	.355	.549	.743
(8, 8)	.000	.001	.009	.032	.100	.214	.405	.595	.786
(8, 9)	.000	.001	.005	.020	.069	.157	.319	.500	.702
(8, 10)	.000	.000	.003	.013	.048	.117	.251	.419	.621
(9, 9)	.000	.000	.003	.012	.044	.109	.238	.399	.601
(9, 10)	.000	.000	.002	.008	.029	.077	.179	.319	.510
(10, 10)	.000	.000	.001	.004	.019	.051	.128	.242	.414

Tabla 10 **(*continúa*)**

	a									
(n_1, n_2)	11	12	13	14	15	16	17	18	19	20
(2, 3)										
(2, 4)										
(2, 5)										
(2, 6)										
(2, 7)										
(2, 8)										
(2, 9)										
(2, 10)										
(3, 3)										
(3, 4)										
(3, 5)										
(3, 6)										
(3, 7)										
(3, 8)										
(3, 9)										
(3, 10)										
(4, 4)										
(4, 5)										
(4, 6)										
(4, 7)										
(4, 8)										
(4, 9)										
(4, 10)										
(5, 5)										
(5, 6)	1.000									
(5, 7)	1.000									
(5, 8)	1.000									
(5, 9)	1.000									
(5, 10)	1.000									
(6, 6)	.998	1.000								
(6, 7)	.992	.999	1.000							
(6, 8)	.984	.998	1.000							
(6, 9)	.972	.994	1.000							
(6, 10)	.958	.990	1.000							
(7, 7)	.975	.996	.999	1.000						
(7, 8)	.949	.988	.998	1.000	1.000					
(7, 9)	.916	.975	.994	.999	1.000					
(7, 10)	.879	.957	.990	.998	1.000					
(8, 8)	.900	.968	.991	.999	1.000	1.000				
(8, 9)	.843	.939	.980	.996	.999	1.000	1.000			
(8, 10)	.782	.903	.964	.990	.998	1.000	1.000			
(9, 9)	.762	.891	.956	.988	.997	1.000	1.000	1.000		
(9, 10)	.681	.834	.923	.974	.992	.999	1.000	1.000	1.000	
(10, 10)	.586	.758	.872	.949	.981	.996	.999	1.000	1.000	1.000

De “Tables for Testing Randomness of Grouping in a Sequence of Alternatives”, C. Eisenhart y F. Swed, *Annals of Mathematical Statistics*, Volumen 14 (1943).

Tabla 11 Valores críticos de coeficiente de correlación de rango de Spearman

n	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
5	0.900	—	—	—
6	0.829	0.886	0.943	—
7	0.714	0.786	0.893	—
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

De "Distribution of Sums of Squares of Rank Differences for Small Samples", E. G. Olds, *Annals of Mathematical Statistics*, Volumen 9 (1938).

Tabla 12 Números aleatorios

Renglón/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93003	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72095	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	53342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953

Tabla 12 (*continúa*)

Renglón/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
29	05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844	69618
31	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
32	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
33	69011	65795	95876	55293	18988	27354	26575	08625	40801	59920	29841	80150	12777	48501
34	25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
36	91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
37	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83517	36103	42791
38	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
39	92157	89634	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509	20468
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
41	98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
42	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
43	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794
44	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
45	76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
46	90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
47	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
48	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
49	95012	68379	93526	70765	10592	04542	76463	54328	02349	17247	28865	14777	62730	92277
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653

Tabla 12 (*continúa*)

Renglón/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
51	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649	91754	72772	02338
52	18629	81953	05520	91962	04739	13092	97662	24822	94730	06496	35090	04822	86774	98289
53	73115	35101	47498	87637	99016	71060	88824	71013	18735	20286	23153	72924	35165	43040
54	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812	12515	98931	91202
55	30405	83946	23792	14422	15059	45799	22716	19792	09983	74353	68668	30429	70735	25499
56	16631	35006	85900	98275	32388	52390	16815	69298	82732	38480	73817	32523	41961	44437
57	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052	91491	60383	19746
58	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318	29686	03387	59846
59	31624	76384	17403	53363	44167	64486	64758	75366	76554	31601	12614	33072	60332	92325
60	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806	08930	85001	87820
61	03931	33309	57047	74211	63445	17361	62825	39908	05607	91284	68833	25570	38818	46920
62	74426	33278	43972	10119	89917	15665	52872	73823	73144	88662	88970	74492	51805	99378
63	09066	00903	20795	95452	92648	45454	09552	88815	16553	51125	79375	97596	16296	66092
64	42238	12426	87025	14267	20979	04508	64535	31355	86064	29472	47689	05974	52468	16834
65	16153	08002	26504	41744	81959	65642	74240	56302	00033	67107	77510	70625	28725	34191
66	21457	40742	29820	96783	29400	21840	15035	34537	33310	06116	95240	15957	16572	06004
67	21581	57802	02050	89728	17937	37621	47075	42080	97403	48626	68995	43805	33386	21597
68	55612	78095	83197	33732	05810	24813	86902	60397	16489	03264	88525	42786	05269	92532
69	44657	66999	99324	51281	84463	60563	79312	93454	68876	25471	93911	25650	12682	73572
70	91340	84979	46949	81973	37949	61023	43997	15263	80644	43942	89203	71795	99533	50501
71	91227	21199	31935	27022	84067	05462	35216	14486	29891	68607	41867	14951	91696	85065
72	50001	38140	66321	19924	72163	09538	12151	06878	91903	18749	34405	56087	82790	70925
73	65390	05224	72958	28609	81406	39147	25549	48542	42627	45233	57202	94617	23772	07896
74	27504	96131	83944	41575	10573	08619	64482	73923	36152	05184	94142	25299	84387	34925
75	37169	94851	39117	89632	00959	16487	65536	49071	39782	17095	02330	74301	00275	48280

Tabla 12 (*continúa*)

Renglón/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
76	11508	70225	51111	38351	19444	66499	71945	05422	13442	78675	84081	66938	93654	59894
77	37449	30362	06694	54690	04052	53115	62757	95348	78662	11163	81651	50245	34971	52924
78	46515	70331	85922	38329	57015	15765	97161	17869	45349	61796	66345	81073	49106	79860
79	30986	81223	42416	58353	21532	30502	32305	86482	05174	07901	54339	58861	74818	46942
80	63798	64995	46583	09785	44160	78128	83991	42885	92520	83531	80377	35909	81250	54238
81	82486	84846	99254	67632	43218	50076	21361	64816	51202	88124	41870	52689	51275	83556
82	21885	32906	92431	09060	64297	51674	64126	62570	26123	05155	59194	52799	28225	85762
83	60336	98782	07408	53458	13564	59089	26445	29789	85205	41001	12535	12133	14645	23541
84	43937	46891	24010	25560	86355	33941	25786	54990	71899	15475	95434	98227	21824	19585
85	97656	63175	89303	16275	07100	92063	21942	18611	47348	20203	18534	03862	78095	50136
86	03299	01221	05418	38982	55758	92237	26759	86367	21216	98442	08303	56613	91511	75928
87	79626	06486	03574	17668	07785	76020	79924	25651	83325	88428	85076	72811	22717	50585
88	85636	68335	47539	03129	65651	11977	02510	26113	99447	68645	34327	15152	55230	93448
89	18039	14367	61337	06177	12143	46609	32989	74014	64708	00533	35398	58408	13261	47908
90	08362	15656	60627	36478	65648	16764	53412	09013	07832	41574	17639	82163	60859	75567
91	79556	29068	04142	16268	15387	12856	66227	38358	22478	73373	88732	09443	82558	05250
92	92608	82674	27072	32534	17075	27698	98204	63863	11951	34648	88022	56148	34925	57031
93	23982	25835	40055	67006	12293	02753	14827	23235	35071	99704	37543	11601	35503	85171
94	09915	96306	05908	97901	28395	14186	00821	80703	70426	75647	76310	88717	37890	40129
95	59037	33300	26695	62247	69927	76123	50842	43834	86654	70959	79725	93872	28117	19233
96	42488	78077	69882	61657	34136	79180	97526	43092	04098	73571	80799	76536	71255	64239
97	46764	86273	63003	93017	31204	36692	40202	35275	57306	55543	53203	18098	47625	88684
98	03237	45430	55417	63282	90816	17349	88298	90183	36600	78406	06216	95787	42579	90730
99	86591	81482	52667	61582	14972	90053	89534	76036	49199	43716	97548	04379	46370	28672
100	38534	01715	94964	87288	65680	43772	39560	12918	86537	62738	19636	51132	25739	56947

Resumen de *Handbook of Tables for Probability and Statistics*, Segunda edición, editado por William H. Beyer (Cleveland: The Chemical Rubber Company, 1968).

RESPUESTAS

Capítulo 1

1.5 **a** $2.45 - 2.65, 2.65 - 2.85$
b $7/30$
c $16/30$

1.9 **a** Aprox. .68
b Aprox. .95
c Aprox. .815
d Aprox. 0

1.13 **a** $\bar{y} = 9.79$; $s = 4.14$
b $k = 1$: $(5.65, 13.93)$; $k = 2$: $(1.51, 18.07)$; $k = 3$: $(-2.63, 22.21)$

1.15 **a** $\bar{y} = 4.39$; $s = 1.87$
b $k = 1$: $(2.52, 6.26)$; $k = 2$: $(0.65, 8.13)$; $k = 3$: $(-1.22, 10)$

1.17 Para el Ejemplo 1.2, rango$/4 = 7.35$; $s = 4.14$; para el Ejemplo 1.3, rango$/4 = 3.04$; $s = 3.17$; para el Ejemplo 1.4, rango$/4 = 2.32$, $s = 1.87$.

1.19 $\bar{y} - s = -19 < 0$

1.21 .84

1.23 **a** 16%
b Aprox. 95%

1.25 **a** 177
c $\bar{y} = 210.8$; $s = 162.17$
d $k = 1$: $(48.6, 373)$; $k = 2$: $(-113.5, 535.1)$; $k = 3$: $(-275.7, 697.3)$

1.27 68% o 231 puntos; 95% o 323 puntos

1.29 .05

1.31 .025

1.33 $(0.5, 10.5)$

1.35 **a** $(172 - 108)/4 = 16$
b $\bar{y} = 136.1$; $s = 17.1$
c $a = 136.1 - 2(17.1) = 101.9$; $b = 136.1 + 2(17.1) = 170.3$

Capítulo 2

2.7 $A = \{\text{dos hombres}\} = \{(M_1, M_2), (M_1, M_3), (M_2, M_3)\}$
$B = \{\text{al menos una mujer}\} = \{(M_1, W_1), (M_2, W_1), (M_3, W_1), (M_1, W_2), (M_2, W_2), (M_3, W_2), (W_1, W_2)\}$
$\bar{B} = \{\text{no hay mujeres}\} = A$; $A \cup B = S$; $A \cap B = \text{nulo}$; $A \cap \bar{B} = A$

2.9 $S = \{A^+, B^+, AB^+, O^+, A^-, B^-, AB^-, O^-\}$

2.11 **a** $P(E_5) = .10$; $P(E_4) = .20$
b $p = .2$

2.13 **a** $E_1 =$ muy probablemente (VL); $E_2 =$ poco probable (SL); $E_3 =$ improbable (U); $E_4 =$ otro (O)
b No; $P(\text{VL}) = .24$, $P(\text{SL}) = .24$, $P(\text{U}) = .40$, $P(\text{O}) = .12$
c .48

2.15 **a** .09
b .19

2.17 **a** .08
b .16
c .14
d .84

2.19 **a** $(V_1, V_1), (V_1, V_2), (V_1, V_3), (V_2, V_1), (V_2, V_2), (V_2, V_3), (V_3, V_1), (V_3, V_2), (V_3, V_3)$
b Es igualmente probable, todos tienen probabilidad de $1/9$.
c $P(A) = 1/3$; $P(B) = 5/9$; $P(A \cup B) = 7/9$; $P(A \cap B) = 1/9$

2.27 **a** $S = \{CC, CR, CL, RC, RR, RL, LC, LR, LL\}$
b $5/9$
c $5/9$

2.29 **c** 1/15
2.31 **a** 3/5; 1/15
b 14/15; 2/5
2.33 **c** 11/16; 3/8; 1/4
2.35 42
2.37 **a** 6! = 720
b .5
2.39 **a** 36
b 1/6
2.41 $9(10)^6$
2.43 504 formas
2.45 408,408
2.49 **a** 8385
b 18,252
c Se requieren 8515
d Sí
2.51 **a** 4/19,600
b 276/19,600
c 4140/19,600
d 15,180/19,600
2.53 **a** 60 puntos muestrales
b 36/60 = .6
2.55 **a** $\binom{90}{10}$
b $\binom{20}{4}\binom{70}{6}\Big/\binom{90}{10} = .111$
2.57 $(4 \times 12)/1326 = .0362$
2.59 **a** .000394
b .00355
2.61 **a** $\dfrac{364^n}{365^n}$
b .5005
2.63 1/56
2.65 5/162
2.67 **a** $P(A) = .0605$
b .001344
c .00029
2.71 **a** 1/3
b 1/5
c 5/7
d 1
e 1/7
2.73 **a** 3/4
b 3/4
c 2/3
2.77 **a** .40 **b** .37 **c** .10
d .67 **e** .6 **f** .33
g .90 **h** .27 **i** .25
2.93 .364
2.95 **a** .1
b .9
c .6
2.97 **a** .999
b .9009
2.101 .05
2.103 **a** .001
b .000125
2.105 .90
2.109 $P(A) \geq .9833$
2.111 .149
2.113 $(.98)^3(.02)$
2.115 $(.75)^4$
2.117 **a** $4(.5)^4 = .25$
b $(.5)^4 = 1/16$
2.119 **a** 1/4
b 1/3
2.121 **a** $1/n$
b $\dfrac{1}{n}; \dfrac{1}{n}$
c $\dfrac{3}{7}$
2.125 1/12
2.127 **a** .857
c No; .8696
d Sí
2.129 .4
2.133 .9412
2.135 **a** .57
b .18
c .3158
d .90
2.137 **a** 2/5
b 3/20
2.139 $P(Y = 0) = (.02)^3$;
$P(Y = 1) = 3(.02)^2(.98)$;
$P(Y = 2) = 3(.02)(.98)^2$;
$P(Y = 3) = (.98)^3$
2.141 $P(Y = 2) = 1/15$; $P(Y = 3) = 2/15$;
$P(Y = 4) = 3/15$; $P(Y = 5) = 4/15$;
$P(Y = 6) = 5/15$
2.145 18!
2.147 .0083
2.149 **a** .4
b .6
c .25
2.151 $4[p^4(1 - p) + p(1 - p)^4]$
2.153 .313
2.155 **a** .5
b .15
c .10
d .875

2.157 .021

2.161 $P(R \leq 3) = 12/66$

2.163 $P(A) = 0.9801$
$P(B) = .9639$

2.165 .916

2.167 $P(Y = 1) = 35/70 = .5$;
$P(Y = 2) = 20/70 = 2/7$;
$P(Y = 3) = 10/70$;
$P(Y = 4) = 4/70$; $P(Y = 5) = 1/70$

2.169 **a** $(4!)^3 = 13{,}824$
b $3456/13{,}824 = .25$

2.173 .25

2.177 **a** .364
b .636
c $(49/50)^n \geq .60$, y n es a lo sumo 25

2.179 **a** $20\left(\frac{1}{2}\right)^6 = .3125$
b $27\left(\frac{1}{2}\right)^{10}$

Capítulo 3

3.1 $P(Y = 0) = .2$, $P(Y = 1) = .7$,
$P(Y = 2) = .1$

3.3 $p(2) = \frac{1}{6}$, $p(3) = \frac{2}{6}$, $p(4) = \frac{1}{2}$

3.5 $p(0) = \frac{2}{6}$, $p(1) = \frac{3}{6}$, $p(3) = \frac{1}{6}$

3.7 $p(0) = \frac{3!}{27} = \frac{6}{27}$, $p(2) = \frac{3}{27}$,
$p(1) = 1 - \frac{6}{27} - \frac{3}{27} = \frac{18}{27}$

3.9 **a** $P(Y = 3) = .000125$,
$P(Y = 2) = .007125$,
$P(Y = 1) = .135375$,
$P(Y = 0) = .857375$
c $P(Y > 1) = .00725$

3.11 $P(X = 0) = \frac{8}{27}$, $P(X = 1) = \frac{12}{27}$,
$P(X = 2) = \frac{6}{27}$, $P(X = 3) = \frac{1}{27}$,
$P(Y = 0) = \frac{2744}{3375}$,
$P(Y = 1) = \frac{588}{3375}$,
$P(Y = 2) = \frac{14}{3375}$,
$P(Y = 3) = \frac{1}{3375}$, $Z = X + Y$,
$P(Z = 0) = \frac{27}{125}$, $P(Z = 1) = \frac{54}{125}$,
$P(Z = 2) = \frac{36}{125}$, $P(Z = 3) = \frac{8}{125}$

3.13 $E(Y) = \frac{1}{4}$, $E(Y^2) = \frac{7}{4}$, $V(Y) = \frac{27}{16}$,
costo $= \frac{1}{4}$

3.15 **a** $P(Y = 0) = .1106$,
$P(Y = 1) = .3594$,
$P(Y = 2) = .3894$,
c $P(Y = 1) = .3594$
d $\mu = E(Y) = 1.56$, $\sigma^2 = .7488$,
$\sigma = 0.8653$
e $(-.1706, 3.2906)$,
$P(0 \leq Y \leq 3) = 1$

3.17 $\mu = E(Y) = .889$,
$\sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2 = .321$
$\sigma = 0.567$, $(\mu - 2\sigma,$
$\mu + 2\sigma) = (-.245, 2.023)$,
$P(0 \leq Y \leq 2) = 1$

3.19 $C = \$85$

3.21 13,800.388

3.23 \$.31

3.25 Firma I: E(utilidad) = \$60,000
E(utilidad total) = \$120,000

3.27 \$510

3.35 .4; .3999

3.39 **a** .1536;
b .9728

3.41 .000

3.43 **a** .1681
b .5282

3.45 P(funciones de alarma) = 0.992

3.49 **a** .151
b .302

3.51 **a** .51775
b .4914

3.53 **a** .0156
b .4219
c 25%

3.57 \$185,000

3.59 \$840

3.61 **a** .672
b .672
c 8

3.67 .07203

3.69 Y es geométrica con $p = .59$

3.73 **a** .009
b .01

3.75 **a** .081
b .81

3.81 2

3.83 $\frac{1}{n}\left(\frac{n-1}{n}\right)^5$

3.87 $E\left(\frac{1}{Y}\right) = -\frac{p\ln(p)}{1-p}$

3.91 \$150; 4500

3.93 **a** .04374
b .99144

3.95 .1

3.97 **a** .128
b .049
c $\mu = 15, \sigma^2 = 60$

3.99 $p(x) = \frac{y!}{(r-1)!(y-r+1)!}p^r q^{y+1-r}$, $y = r-1, r, r+1, \ldots$

3.101 **a** $\frac{5}{11}$
b $\frac{r}{y_0}$

3.103 $\frac{1}{42}$

3.105 **b** .7143
c $\mu = 1.875$, $\sigma = .7087$

3.107 hipergeométrica con $N = 6, n = 2$, y $r = 4$.

3.109 **a** .0238
b .9762
c .9762

3.111 **a** $p(0) = \frac{14}{30}, p(1) = \frac{14}{30}$, $p(2) = \frac{2}{30}$
b $p(0) = \frac{5}{30}, p(1) = \frac{15}{30}$, $p(2) = \frac{9}{30}, p(3) = \frac{1}{30}$

3.113 $P(Y \leq 1) = .187$

3.115 $p(0) = \frac{1}{5}, p(1) = \frac{3}{5}, p(2) = \frac{1}{5}$

3.117 **a** $P(Y = 0) = .553$
b $E(T) = 9.5, V(T) = 28.755$, $\sigma = 5.362$

3.119 .016

3.121 **a** .090
b .143
c .857
d .241

3.123 .1839

3.125 $E(S) = 7, V(S) = 700$; no

3.127 .6288

3.129 23 segundos

3.131 .5578

3.133 .1745

3.135 .9524

3.137 .1512

3.139 40

3.141 \$1300

3.149 Binomial, $n = 3$ y $p = .6$

3.151 Binomial, $n = 10$ y $p = .7$, $P(Y \leq 5) = .1503$

3.153 **a** Binomial, $n = 5$ y $p = .1$
b Geométrica, $p = \frac{1}{2}$
c Poisson, $\lambda = 2$

3.155 **a** $E(Y) = \frac{7}{3}$
b $V(Y) = \frac{5}{9}$
c $p(1) = \frac{1}{6}, p(2) = \frac{2}{6}, p(3) = \frac{3}{6}$

3.167 **a** .64
b $C = 10$

3.169 **d** $p(-1) = 1/(2k^2)$, $p(0) = 1 - (1/k^2), p(1) = 1(2k^2)$

3.171 (85, 115)

3.173 **a** $p(0) = \frac{1}{8}, p(1) = \frac{3}{8}, p(2) = \frac{3}{8}$, $p(3) = \frac{1}{8}$
c $E(Y) = 1.5, V(Y) = .75$, $\sigma = .866$

3.175 **a** 38.4
b 5.11

3.177 (61.03, 98.97)

3.179 No, $P(Y \geq 350) \leq \frac{1}{(2.98)^2} = .1126.$

3.181

	p = Fracción defectuosa	P(aceptación)
a	0	1
b	.10	.5905
c	.30	.1681
d	.50	.0312
e	1.0	0

3.185 **a** .2277
b No improbable

3.187 **a** .023
b 1.2
c \$1.25

3.189 $1 - (.99999)^{10,000}$

3.191 $V(Y) = .4$

3.193 .476

3.195 **a** .982
b $P(W \geq 1) = 1 - e^{-12}$

3.197 **a** .9997
b $n = 2$

3.199 **a** .300
b .037

3.201 (18.35, 181.65)

3.203 **a** $E[Y(t)] = k(e^{2\lambda t} - e^{\lambda t})$
b 3.2974, 2.139

3.205 .00722

3.207 **a** $p(2) = .084$, $P(Y \leq 2) = .125$
b $P(Y > 10) = .014$

3.209 .0837

3.211 3

3.213 **a** .1192
b .117

3.215 **a** $n[1 + k(1 - .95^k)]$
b $g(k)$ se minimiza en $k = 5$ y $g(5) = .4262$.
c $.5738N$

Capítulo 4

4.7 **a** $P(2 \leq Y < 5) = 0.591$, $P(2 < Y < 5) = .289$, y no es igual
b $P(2 \leq Y \leq 5) = 0.618$, $P(2 < Y \leq 5) = 0.316$, y no es igual
c Y no es una variable aleatoria continua, de modo que no se cumplen los resultados anteriores

4.9 **a** Y es una variable aleatoria discreta
b Estos valores son 2, 2.5, 4, 5.5, 6, y 7.
c $p(2) = \frac{1}{8}$, $p(2.5) = \frac{1}{16}$, $p(4) = \frac{5}{16}$, $p(5.5) = \frac{1}{8}$, $p(6) = \frac{1}{16}$, $p(7) = \frac{5}{16}$
d $\phi_{.5} = 4$

4.11 **a** $c = \frac{1}{2}$
b $F(y) = \frac{y^2}{4}$, $0 \leq y \leq 2$
d .75
e .75

4.13 **a**
$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 \leq y \leq 1 \\ y - \frac{1}{2} & 1 < y \leq 1.5 \\ 1 & y > 1.5 \end{cases}$$
b .125
c .575

4.15 **a** Para $b \geq 0$, $f(y) \geq 0$; también, $\int_{-q}^{q} f(y) = 1$
b $F(y) = 1 - \frac{b}{y}$, para $y \geq b$; 0 en otra parte.
c $\frac{b}{(b + c)}$
d $\frac{(b + c)}{(b + d)}$

4.17 **a** $c = \frac{3}{2}$
b $F(y) = \frac{y^3}{2} + \frac{y^2}{2}$, para $0 \leq y \leq 1$

d $F(-1) = 0, F(0) = 0, F(1) = 1$

e $\frac{3}{16}$

f $\frac{104}{123}$

4.19 **a** $f(y) = \begin{cases} 0 & y \le 0 \\ .125 & 0 < y < 2 \\ .125y & 2 \le y < 4 \\ 0 & y \ge 4 \end{cases}$

b $\frac{7}{16}$

c $\frac{13}{16}$

d $\frac{7}{9}$

4.21 $E(Y) = .708, V(Y) = .0487$

4.25 $E(Y) = 31/12, V(Y) = 1.160$

4.27 $4.65, .012

4.29 $E(Y) = 60, V(Y) = \frac{1}{3}$

4.31 $E(Y) = 4$

4.33 **a** $E(Y) = 5.5, V(Y) = .15$

b Usando el teorema de Tchebysheff, el intervalo es (5, 6.275).

c Sí; $P(Y) = .5781$

4.37 $E(Y) = 0$

4.39 .5; .25

4.45 **a** $P(Y < 22) = \frac{2}{5} = .4$

b $P(Y > 24) = \frac{1}{5} = .2$

4.47 **a** $P(Y > 2) = \frac{3}{4}$

b $c_0 + c_1\left[\frac{4}{3} + 9\right]$

4.49 $\frac{3}{4}$

4.51 $\frac{1}{3}$

4.53 **a** $\frac{1}{8}$

b $\frac{1}{8}$

c $\frac{1}{4}$

4.55 **a** $\frac{2}{7}$

b $\mu = .015, V(Y) = .00041$

4.57 $E\left(\frac{\pi}{6}D^3\right) = .0000065\pi,$

$V\left(\frac{\pi}{6}D^3\right) = .0003525\pi^2$

4.59 **a** $z_0 = 0$

b $z_0 = 1.10$

c $z_0 = 1.645$

d $z_0 = 2.576$

4.63 **a** $P(Z > 1) = .1587$

b Se obtiene la misma respuesta.

4.65 $425.60

4.67 $\mu = 3.000$ in.

4.69 .2660

4.71 **a** .9544

b .8297

4.73 **a** .406

b 960.5 mm

4.75 $\mu = 7.301$

4.77 **a** 0.758

b 22.2

4.87 **a** $\phi_{.05} = .70369.$

b $\phi_{.05} = .35185$

4.89 **a** $\beta = .8$

b $P(Y \le 1.7) = .8806$

4.91 **a** .1353

b 460.52 cfs

4.93 **a** .5057

b 1936

4.97 .3679

4.99 **a** .7358

4.101 **a** $E(Y) = 1.92$

b $P(Y > 3) = .21036$

d $P(2 \le Y \le 3) = .12943$

4.103 $E(A) = 200\pi, V(A) = 200{,}000\pi^2$

4.105 **a** $E(Y) = 3.2, V(Y) = 6.4$

b $P(Y > 4) = .28955$

4.107 **a** (0, 9.657), porque Y debe ser positiva.

b $P(Y < 9.657) = .95338$

4.109 $E(L) = 276$, $\mathrm{V}(L) = 47{,}664$

4.111 **d** $\sqrt{\beta}\Gamma\left(\alpha + \frac{1}{2}\right) / \Gamma(\alpha)$ si $\alpha > 0$

e $\frac{1}{\beta(\alpha - 1)}$ si $\alpha > 1$, $\frac{\Gamma(\alpha - \frac{1}{2})}{\sqrt{\beta}\Gamma(\alpha)}$ si $\alpha > \frac{1}{2}$, $\frac{1}{\beta^2(\alpha - 1)(\alpha - 2)}$ si $\alpha > 2$

4.123 **a** $k = 60$

b $\phi_{.95} = 0.84684$

4.125 $E(Y) = \frac{3}{5}, V(Y) = \frac{1}{25}$

4.129 $E(C) = \frac{52}{3}, V(C) = 29.96$

4.131 **a** .75

b .2357

4.133 **a** $c = 105$

b $\mu = \frac{3}{8}$
c $\sigma = .1614$
d .02972

4.139 $m_X(t) = \exp\{t(4-3\mu) + (1/2)(9\sigma^2 t^2)\}$ normal, $E(X) = 4 - 3\mu$, $V(X) = 9\sigma^2$, unicidad de funciones de generación de momento

4.141 $m(t) = \dfrac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$

4.143 $\alpha\beta$, $\alpha\beta^2$

4.145 **a** $\frac{2}{5}$
b $\dfrac{1}{(t+1)}$
c 1

4.147 $\sigma = \frac{1}{2}$

4.149 1

4.151 El valor de 2000 es .53 desviación estándar sobre la media. Así, esperaríamos que C rebase 2000 con frecuencia.

4.153 (6.38, 28.28)

4.155 \$113.33

4.157 **a** $F(x) =$
$$\begin{cases} 0, & x < 0 \\ (1/100)e^{-x/100}, & 0 \le x < 200 \\ 1, & x \ge 200 \end{cases}$$
b 86.47

4.159 **a** $F_1(y) =$
$$\begin{cases} 0 & y < 0 \\ \dfrac{.1}{.1 + .15} = .4 & 0 \le y < 5; \\ 1 & y \ge .5 \end{cases}$$
$F_2(y) =$
$$\begin{cases} 0 & y < 0 \\ 4y^2/3 & 0 \le y < .5 \\ (4y-1)/3 & .5 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$
b $F(y) = 0.25F_1(y) + 0.75F_2(y)$
c $E(Y) = .533$, $V(Y) = .076$

4.161 $\phi_{.9} = 85.36$

4.163 $1 - (.927)^5 = .3155$

4.165 **a** $c = 4$
b $E(Y) = 1$, $V(Y) = .5$
c $m(t) = \dfrac{1}{(1 - .5t)^2}$, $t < 2$

4.167 $E(Y^k) = \dfrac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}$

4.169 $e^{-2.5} = .082$

4.171 **a** $E(W) = \frac{1}{2}$, $V(W) = \frac{1}{4}$
b $1 - e^{-6}$

4.173 $f(r) = 2\lambda\pi r e^{-\lambda\pi r^2}$, $r > 0$

4.175 $\sqrt{2} = 1.414$

4.179 $k = (.4)^{1/3} = .7368$

4.181 $m(t) = \exp(t^2/2)$; 0; 1

4.183 **a** $E(Y) = 598.74$ g
$V(Y) = e^{22}(e^{16} - 1)10^{-4}$
b (0, 3,570,236.1)
c .8020

4.187 **a** $e^{-2.5} = .082$
b .0186

4.189 $E(Y) = 0$. También, es claro que $V(Y) = E(Y^2) = \dfrac{1}{n-1}$.

4.191 **c** $1 - e^{-4}$

4.193 150

4.195 **a** 12
b $w = 120$

Capítulo 5

5.1 **a**

		y_1 = 0	1	2
y_2	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$
	1	$\frac{2}{9}$	$\frac{2}{9}$	0
	2	$\frac{1}{9}$	0	0

b $F(1, 0) = \frac{1}{3}$

5.3 $\dfrac{\binom{4}{y_1}\binom{3}{y_2}\binom{2}{3 - y_1 - y_2}}{\binom{9}{3}}$, donde $0 \le y_1$, $0 \le y_2$, y $y_1 + y_2 \le 3$.

5.5 **a** .1065
b .5

5.7 **a** .00426
b .8009

5.9 **a** $k = 6$
b $\frac{31}{64}$

5.11 **a** $\frac{29}{32}$
b $\frac{1}{4}$

5.13 **a** $F\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{9}{16}$

b $F\left(\frac{1}{2}, 2\right) = \frac{13}{16}$

c .65625

5.15 **a** $e^{-1} - 2e^{-2}$

b $\frac{1}{2}$

c e^{-1}

5.17 .50

5.19 **a**

y_1	0	1	2
$p_1(y_1)$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

b No

5.21 **a** Hipergeométrica con $N = 9$, $n = 3$, y $r = 4$.

b $\frac{2}{3}$

c $\frac{8}{15}$

5.23 **a** $f_2(y_2) = \frac{3}{2} - \frac{3}{2}y_2^2,\ 0 \le y_2 \le 1$

b Definida sobre $y_2 \le y_1 \le 1$ si $y_2 \ge 0$

c $\frac{1}{3}$

5.25 **a** $f_1(y_1) = e^{-y_1},\ y_1 > 0;$ $f_2(y_2) = e^{-y_2},\ y_2 > 0$

b $P(1 < Y_1 < 2.5) = P(1 < Y_2 < 2.5) = e^{-1} - e^{-2.5} = .2858$

c $y_2 > 0$

d $f(y_1|y_2) = f_1(y_1) = e^{-y_1},\ y_1 > 0$

e $f(y_2|y_1) = f_2(y_2) = e^{-y_2},\ y_2 > 0$

f igual

g igual

5.27 **a** $f_1(y_1) = 3(1 - y_1)^2,\ 0 \le y_1 \le 1;$ $f_2(y_2) = 6y_2(1 - y_2),\ 0 \le y_2 \le 1$

b $\frac{32}{63}$

c $f(y_1|y_2) = \frac{1}{y_2},\ 0 \le y_1 \le y_2,$ si $y_2 \le 1$

d $f(y_2|y_1) = \frac{2(1 - y_2)}{(1 - y_1)^2},$ $y_1 \le y_2 \le 1$ si $y_1 \ge 0$

e $\frac{1}{4}$

5.29 **a** $f_2(y_2) = 2(1 - y_2),\ 0 \le y_2 \le 1;$ $f_1(y_1) = 1 - |y_1|$, para $-1 \le y_1 \le 1$

b $\frac{1}{3}$

5.31 **a** $f_1(y_1) = 20y_1(1 - y_1)^2,\ 0 \le y_1 \le 1$

b $f_2(y_2) = \begin{cases} 15(1 + y_2)^2 y_2^2, & -1 \le y_2 < 0 \\ 15(1 - y_2)^2 y_2^2, & 0 \le y_2 \le 1 \end{cases}$

c $f(y_2|y_1) = \frac{3}{2}y_2^2(1 - y_1)^{-3},$ para $y_1 - 1 \le y_2 \le 1 - y_1$

d .5

5.33 **a** $f_1(y_1) = y_1 e^{-y_1},\ y_1 \ge 0;$ $f_2(y_2) = e^{-y_2},\ y_2 \ge 0$

b $f(y_1|y_2) = e^{-(y_1 - y_2)},\ y_1 \ge y_2$

c $f(y_2|y_1) = 1/y_1,\ 0 \le y_2 \le y_1$

5.35 .5

5.37 e^{-1}

5.41 $\frac{1}{4}$

5.45 No

5.47 Dependiente

5.51 **a** $f(y_1, y_2) = f_1(y_1) f_2(y_2)$ de modo que Y_1 y Y_2 son independientes.

b Sí, las probabilidades condicionales son iguales que las probabilidades marginales.

5.53 No, son dependientes.

5.55 No, son dependientes.

5.57 No, son dependientes.

5.59 No, son dependientes.

5.61 Sí, son independientes.

5.63 $\frac{1}{4}$

5.65 Exponencial, media 1

5.69 **a** $f(y_1, y_2) = \left(\frac{1}{9}\right) e^{-(y_1 + y_2)/3},$ $y_1 > 0,\ y_2 > 0$

b $P(Y_1 + Y_2 \le 1) = 1 - \frac{4}{3}e^{-1/3} = .0446$

5.71 **a** $\frac{1}{4}$

b $\frac{23}{144}$

5.73 $\frac{4}{3}$

5.75 **a** 2

b .0249

c .0249

d 2

e Son iguales.

5.77 **a** $\frac{1}{4}; \frac{1}{2}$

b $E(Y_1^2) = 1/10,\ V(Y_1) = \frac{3}{80},$ $E(Y_2^2) = \frac{3}{10},\ V(Y_2) = \frac{1}{20}$

c $-\frac{5}{4}$

5.79 0

5.81 1

5.83 1

5.85 **a** $E(Y_1) = E(Y_2) = 1$ (ambas distribuciones marginales son exponenciales con media 1)
b $V(Y_1) = V(Y_2) = 1$
c $E(Y_1 - Y_2) = 0$
d $E(Y_1Y_2) = 1 - \frac{\alpha}{4}$, así $\text{Cov}(Y_1, Y_2) = -\frac{\alpha}{4}$
e $\left(-2\sqrt{2 + \frac{\alpha}{2}},\ 2\sqrt{2 + \frac{\alpha}{2}}\right)$

5.87 **a** $E(Y_1 + Y_2) = \nu_1 + \nu_2$
b $V(Y_1 + Y_2) = 2\nu_1 + 2\nu_2$

5.89 $\text{Cov}(Y_1, Y_2) = -\frac{2}{9}$. Cuando aumenta el valor de Y_1 el de Y_2 tiende a disminuir.

5.91 $\text{Cov}(Y_1, Y_2) = 0$

5.93 **a** 0
b Dependiente
c 0
d No necesariamente independiente

5.95 Las distribuciones marginales para Y_1 y Y_2 son

y_1	-1	0	1	y_2	0	1
$p_1(y_1)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$p_2(y_2)$	$\frac{2}{3}$	$\frac{1}{3}$

$\text{Cov}(Y_1, Y_2) = 0$

5.97 **a** 2
b Imposible
c 4 (una asociación lineal positiva perfecta)
d -4 una asociación lineal negativa perfecta)

5.99 0

5.101 **a** $-\frac{\alpha}{4}$

5.103 $E(3Y_1 + 4Y_2 - 6Y_3) = -22$, $V(3Y_1 + 4Y_2 - 6Y_3) = 480$

5.105 $\frac{1}{9}$

5.107 $E(Y_1 + Y_2) = 2/3$ y $V(Y_1 + Y_2) = \frac{1}{18}$

5.109 (11.48, 52.68)

5.111 $E(G) = 42$, $V(G) = 25$; el valor \$70 es $\frac{70 - 42}{5} = 7.2$ desviaciones estándar arriba de la media, un valor improbable.

5.115 **b** $V(Y) = 38.99$
c El intervalo es14 .7 $\pm 2\sqrt{38.99}$ o bien (0, 27.188)

5.117 $p_1 - p_2$, $\frac{N - n}{n(N - 1)}[p_1 + p_2 - (p_1 - p_2)^2]$

5.119 **a** .0823
b $E(Y_1) = \frac{n}{3}$, $V(Y_1) = \frac{2n}{9}$
c $\text{Cov}(Y_2, Y_3) = -\frac{n}{9}$
d $E(Y_2 - Y_3) = 0$, $V(Y_2 - Y_3) = \frac{2n}{3}$

5.121 **a** .0972
b .2; .072

5.123 .08953

5.125 **a** .046
b .2262

5.127 **a** .2759
b .8031

5.133 **a** $\frac{y_2}{2}$
b $\frac{1}{4}$

5.135 **a** $\frac{3}{2}$
b 1.25

5.137 $\frac{3}{8}$

5.139 **a** $n\alpha\beta$
b $\lambda\alpha\beta$

5.141 $E(Y_2) = \frac{\lambda}{2}$, $V(Y_2) = \frac{2\lambda^2}{3}$

5.143 $m_U(t) = (1 - t^2)^{-1/2}$, $E(U) = 0$, $V(U) = 1$

5.145 $\frac{1}{3}$

5.147 $\frac{11}{36}$

5.149 **a** $f(y_1) = 3y_1^2,\ 0 \le y_1 \le 1$
$f(y_2) = \frac{3}{2}(1 - y_2^2),\ 0 \le y_2 \le 1$
b $\frac{23}{44}$
c $f(y_1|y_2) = \frac{2y_1}{(1 - y_2^2)},\ y_2 \le y_1 \le 1$
d $\frac{5}{12}$

5.157 $p(y) = \binom{y + \alpha - 1}{y}\left(\frac{\beta}{\beta + 1}\right)^y \left(\frac{1}{\beta + 1}\right)^\alpha$, $y = 0, 1, 2, \ldots$

5.161 $E(\bar{Y} - \bar{X}) = \mu_1 - \mu_2$, $V(\bar{Y} - \bar{X}) = \sigma_1^2/n + \sigma_2^2/m$

5.163 **b** $F(y_1, y_2) = y_1 y_2[1 - \alpha(1 - y_1)(1 - y_2)]$
c $f(y_1, y_2) = 1 - \alpha[(1 - 2y_1)(1 - 2y_2)]$, $0 \le y_1 \le 1, 0 \le y_2 \le 1$
d Seleccione dos valores diferentes para α con $-1 \le \alpha \le 1$.

5.165 **a** $(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^n$
b $m(t, 0, 0)$
c $\text{Cov}(X_1, X_2) = -np_1p_2$

Capítulo 6

6.1 **a** $\frac{1-u}{2}$, $-1 \le u \le 1$
b $\frac{u+1}{2}$, $-1 \le u \le 1$
c $\frac{1}{\sqrt{u}} - 1$, $0 \le u \le 1$
d $E(U_1) = -1/3$, $E(U_2) = 1/3$, $E(U_3) = 1/6$
e $E(2Y - 1) = -1/3$, $E(1 - 2Y) = 1/3$, $E(Y^2) = 1/6$

6.3 **b** $f_U(u) = \begin{cases} (u+4)/100, & -4 \le u \le 6 \\ 1/10, & 6 < u \le 11 \end{cases}$
c 5.5833

6.5 $f_U(u) = \frac{1}{16}\left(\frac{u-3}{2}\right)^{-1/2}$, $5 \le u \ge 53$

6.7 **a** $f_U(u) = \frac{1}{\sqrt{\mu}\sqrt{2}} u^{-1/2} e^{-u/2}$, $u \ge 0$
b U tiene una distribución gamma con $\alpha = 1/2$ y $\beta = 2$ (recuerde que $\Gamma(1/2) = \sqrt{\mu}$).

6.9 **a** $f_U(u) = 2u, 0 \le u \le 1$
b $E(U) = 2/3$
c $E(Y_1 + Y_2) = 2/3$

6.11 **a** $f_U(u) = 4ue^{-2u}$, $u \ge 0$, una densidad gamma con $\alpha = 2$ y $\beta = 1/2$
b $E(U) = 1$, $V(U) = 1/2$

6.13 $f_U(u) = F'_U(u) = \frac{u}{\beta^2} e^{-u/\beta}$, $u > 0$

6.15 $[-\ln(1 - U)]^{1/2}$

6.17 **a** $f(y) = \frac{\alpha y^{\alpha-1}}{\theta^\alpha}$, $0 \le y \le \theta$
b $Y = \theta U^{1/\alpha}$
c $y = 4\sqrt{u}$. Los valores son 2.0785, 3.229, 1.5036, 1.5610, 2.403.

6.25 $f_U(u) = 4ue^{-2u}$ para $u \ge 0$

6.27 **a** $f_Y(y) = \frac{2}{\beta} we^{-w^2/\beta}$, $w \ge 0$, que es densidad Weibull con $m = 2$.
b $E(Y^{k/2}) = \Gamma\left(\frac{k}{2} + 1\right)\beta^{k/2}$

6.29 **a** $f_W(w) = \frac{1}{\Gamma\left(\frac{3}{2}\right)(kT)^{3/2}} w^{1/2} e^{-w/kT} \quad w > 0$
b $E(W) = \frac{3}{2}kT$

6.31 $f_U(u) = \frac{2}{(1+u)^3}$, $u \ge 0$

6.33 $f_U(u) = 4(80 - 31u + 3u^2)$, $4.5 \le u \le 5$

6.35 $f_U(u) = -\ln(u)$, $0 \le u \le 1$

6.37 **a** $m_{Y_1}(t) = 1 - p + pe^t$
b $m_W(t) = E(e^{tW}) = [1 - p + pe^t]^n$

6.39 $f_U(u) = 4ue^{-2u}$, $u \ge 0$

6.43 **a** $\bar{Y}$ tiene una distribución normal con media μ y varianza σ^2/n
b $P(|\bar{Y} - \mu| \le 1) = .7888$
c Las probabilidades son .8664, .9544, .9756. Por lo tanto, cuando aumenta el tamaño muestral, aumenta la probabilidad de que $P(|\bar{Y} - \mu| \le 1)$

6.45 $c = \$190.27$

6.47 $P(U > 16.0128) = .025$

6.51 La distribución de $Y_1 + (n_2 - Y_2)$ es binomial con $n_1 + n_2$ intentos y probabilidad de éxito $p = .2$

6.53 **a** Binomial (nm, p) donde $n_i = m$
b Binomial $(n_1 = n_2 + \ldots n_n, p)$
c Hipergeométrico $(r = n, N = n_1 + n_2 + \ldots n_n)$

6.55 $P(Y \ge 20) = .077$

6.65 **a** $f(u_1, u_2) = \frac{1}{2\mu} e^{-[u_1^2 + (u_2 - u_1)^2]/2} = \frac{1}{2\mu} e^{-\left(2u_1^2 - 2u_1u_2 + u_2^2\right)/2}$
b $E(U_1) = E(Z_1) = 0$, $E(U_2) = E(Z_1 + Z_2) = 0$, $V(U_1) = V(Z_1) = 1$, $V(U_2) = V(Z_1 + Z_2) = V(Z_1) + V(Z_2) = 2$, $\text{Cov}(U_1, U_2) = E(Z_1^2) = 1$

c No independiente porque $\rho \neq 0$.
d Ésta es la distribución normal bivariada con $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = 1, \sigma_2^2 = 2$, y $\rho = \dfrac{1}{\sqrt{2}}$

6.69 **a** $f(y_1, y_2) = \dfrac{1}{y_1^2 y_2^2}$, $y_1 > 1$, $y_2 > 1$
e No

6.73 **a** $g_{(2)}(u) = 2u, 0 \le u \le 1$
b $E(U_2) = 2/3, V(U_2) = 1/18$

6.75 $(10/15)^5$

6.77 **a** $\dfrac{n!}{(j-1)!(k-1-j)!(n-k)!} \dfrac{y_j^{j-1}[y_k - y_j]^{k-1-j}[\theta - y_k]^{n-k}}{\theta^n}$, $0 \le y_j < y_k \le \theta$
b $\dfrac{(n-k+1)j}{(n+1)^2(n+2)}\theta^2$
c $\dfrac{(n-k+j+1)(k-j)}{(n+1)^2(n+2)}\theta^2$

6.81 **b** $1 - e^{-9}$

6.83 $1 - (.5)^n$

6.85 .5

6.87 **a** $g_{(1)}(y) = e^{-(y-4)}, y \ge 4$
b $E(Y_{(1)}) = 5$

6.89 $f_R(r) = n(n-1)r^{n-2}(1-r)$, $0 \le r \le 1$

6.93 $f(w) = \dfrac{2}{3}\left(\dfrac{1}{\sqrt{w}} - w\right), 0 \le w \le 1$

6.95 **a** $f_{U_1}(u) = \begin{cases} \dfrac{1}{2} & 0 \le u \le 1 \\ \dfrac{1}{2u^2} & u > 1 \end{cases}$
b $f_{U_2}(u) = ue^{-u}, 0 \le u$
c Igual que el Ej. 6.35.

6.97 $p(W = 0) = p(0) = .0512$, $p(1) = .2048$, $p(2) = .3264$, $p(3) = .2656$, $p(4) = .1186$, $p(5) = .0294$, $p(6) = .0038$, $p(7) = .0002$

6.101 $f_U(u) = 1, 0 \le u \le 1$ Por lo tanto, U tiene una distribución uniforme en (0, 1)

6.103 $\dfrac{1}{\pi(1 + u_1^2)}, \infty < u_1 < \infty$

6.105 $\dfrac{1}{B(\alpha, \beta)} u^{\beta-1}(1-u)^{\alpha-1}, 0 < u < 1$

6.107 $f_U(u) = \begin{cases} \dfrac{1}{4\sqrt{u}} & 0 \le u < 1 \\ \dfrac{1}{8\sqrt{u}} & 1 \le u \le 9 \end{cases}$

6.109 $P(U = C_1 - C_3) = .4156$; $P(U = C_2 - C_3) = .5844$

Capítulo 7

7.9 **a** .7698
b Para $n = 25, 36, 69$, y 64, las probabilidades son (respectivamente) .8664, .9284, .9642, .9836.
c Las probabilidades aumentan con n.
d Sí

7.11 .8664

7.13 .9876

7.15 **a** $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$
b $V(\bar{X} - \bar{Y}) = \sigma_1^2/m + \sigma_2^2/n$
c Los dos tamaños muestrales deberían ser al menos 18.

7.17 $P\left(\sum_{i=1}^{6} Z_i^2 \le 6\right) = .57681$

7.19 $P(S^2 \ge .065) = .10$

7.21 **a** $b = 2.42$
b $a = .656$
c .95

7.27 **a** .17271
b .23041
d .40312

7.31 **a** 5.99, 4.89, 4.02, 3.65, 3.48, 3.32
c 13.2767
d $13.2767/3.32 \approx 4$

7.35 **a** $E(F) = 1.029$
b $V(F) = .076$
c 3 está 7.15 desviaciones estándar arriba de esta media; valor improbable.

7.39 **a** normal, $E(\hat{\theta}) = \theta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$
$V(\hat{\theta}) = \left(\dfrac{c_1^2}{n_1} + \dfrac{c_2^2}{n_2} + \cdots + \dfrac{c_k^2}{n_k}\right)\sigma^2$
b χ^2 con $n_1 + n_2 + \cdots + n_k - k$ gl
c t con $n_1 + n_2 + \cdots + n_k - k$ gl

7.43 .9544

7.45 .0548

7.47 153

7.49 .0217

7.51 664

7.53 **b** $\bar{Y}$ es aproximadamente normal: .0132.

7.55 **a** muestra aleatoria; aproximadamente 1.
b .1271

7.57 .0062
7.59 .0062
7.61 $n = 51$
7.63 56 clientes
7.65 **a** Exacta: .91854; aproximación normal: .86396.
7.67 **a** $n = 5$ (exacta: .99968; aproximada: .95319); $n = 10$ (exacta: .99363; aproximada: .97312); $n = 15$ (exacta: .98194; aproximada: .97613); $n = 20$ (exacta: .96786; aproximada: .96886)
7.71 **a** $n > 9$
b $n > 14$, $n > 14$, $n > 36$, $n > 36$, $n > 891$, $n > 8991$
7.73 .8980
7.75 .7698
7.77 61 clientes
7.79 **a** Usando la aproximación normal: .7486.
b Usando la probabilidad binomial exacta: .729.
7.81 **a** .5948
b Con $p = .2$ y .3, las probabilidades son .0559 y.0017 respectivamente.
7.83 **a** .36897
b .48679
7.85 .8414
7.87 .0041
7.89 $\mu = 10.15$
7.91 Como X, Y, y W están normalmente distribuidas, así están $\bar{X}$, $\bar{Y}$, y $\bar{W}$.

$$\mu_U = E(U) = .4\mu_1 + .2\mu_2 + .4\mu_3$$
$$\sigma_U^2 = V(U) = .16\left(\frac{\sigma_1^2}{n_1}\right) + .04\left(\frac{\sigma_2^2}{n_2}\right) + .16\left(\frac{\sigma_3^2}{n_3}\right)$$

7.95 **a** F con numerador. gl = 1, denominador. gl = 9
b F con numerador. gl = 9, denominador. gl = 1
c $c = 49.04$
7.97 **b** .1587
7.101 .8413
7.103 .1587
7.105 .264

Capítulo 8

8.3 **a** $B(\hat{\theta}) = a\theta + b - \theta = (a-1)\theta + b$
b Sea $\hat{\theta}^* = (\hat{\theta} - b)/a$
8.5 **a** $\text{MSE}(\hat{\theta}^*) = V(\hat{\theta}^*) = V(\hat{\theta})/a^2$
8.7 $a = \dfrac{\sigma_2^2 - c}{\sigma_1^2 + \sigma_2^2 - 2c}$
8.9 $\bar{Y} - 1$
8.11 $\hat{\theta}_3 - 9\hat{\theta}_2 + 54$
8.13 **b** $[n^2/(n-1)](Y/n)[1 - (Y/n)]$
8.15 **a** $\left(\dfrac{1}{3n-1}\right)\beta$
b $\text{MSE}(\hat{\beta}) = \dfrac{2}{(3n-1)(3n-2)}\beta^2$
8.17 **a** $(1-2p)/(n+2)$
b $\dfrac{np(1-p) + (1-2p)^2}{(n+2)^2}$
c p será cercana a .5.
8.19 $\text{MSE}(\hat{\theta}) = \beta^2$
8.21 $11.5 \pm .99$
8.23 **a** 11.3 ± 1.54
b 1.3 ± 1.7
c $.17 \pm .08$
8.25 **a** $-.7$
b .404
8.27 **a** $.601 \pm .031$
8.29 **a** $-.06 \pm .045$
8.31 **a** $-.03 \pm .041$
8.33 $.7 \pm .205$
8.35 **a** 20 ± 1.265
b -3 ± 1.855, sí
8.37 1020 ± 645.1
8.39 $\left(\dfrac{2Y}{9.48773}, \dfrac{2Y}{.71072}\right)$
8.41 **a** $(Y^2/5.02389, Y^2/.0009821)$
b $Y^2/.0039321$
c $Y^2/3.84146$
8.43 **b** $[Y_{(n)}](.95)^{-1/n}$
8.45 **a** $Y/.05132$
b 80%
8.47 **c** (2.557, 11.864)
8.49 **c** (3.108, 6.785)
8.57 $.51 \pm .04$
8.59 **a** $.78 \pm .021$
8.61 (15.46, 36.94)
8.63 **a** $.78 \pm .026$ o (.754, .806)
8.65 **a** $.06 \pm .117$ o $(-.057, .177)$
8.67 **a** $7.2 \pm .751$
b $2.5 \pm .738$
8.69 $.22 \pm .34$ o $(-.12, .56)$
8.71 $n = 100$

8.73 $n = 2847$
8.75 $n = 136$
8.77 $n = 497$
8.79 **a** $n = 2998$
b $n = 1618$
8.81 60.8 ± 5.701
8.83 **a** 3.4 ± 3.7
b $.7 \pm 3.32$
8.85 -1 ± 4.72
8.87 $(-.624, .122)$
8.91 $(-84.39, -28.93)$
8.93 **a** $2\bar{X} + \bar{Y} \pm 1.96\sigma\sqrt{\frac{4}{n} + \frac{3}{m}}$
b $2\bar{X} + \bar{Y} \pm t_{\alpha/2}S\sqrt{\frac{4}{n} + \frac{3}{m}}$, donde
$$S^2 = \frac{\sum(Y_i - \bar{Y})^2 + 1/3\sum(X_i - \bar{X})^2}{n + m - 2}$$
8.95 $(.227, 2.196)$
8.99 **a** $\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha}}}$
b $\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha}}}$
8.101 $s^2 = .0286$; $(.013, .125)$
8.103 $(1.407, 31.264)$; no
8.105 $1 - 2(.0207) = .9586$
8.107 765 semillas
8.109 **a** $.0625 \pm .0237$
b 563
8.111 $n = 38{,}416$
8.113 $n = 768$
8.115 $(29.30, 391.15)$
8.117 11.3 ± 1.44
8.119 3 ± 3.63
8.121 $-.75 \pm .77$
8.123 $.832 \pm .015$
8.125 **a** $\frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2}$
b $\left(\frac{S_2^2}{S_1^2 F_{\nu_2,\nu_1,\alpha/2}}, \frac{S_2^2}{S_1^2} F_{\nu_1,\nu_2,\alpha/2}\right)$
$\nu_i = n_i - 1, i = 1, 2$
8.129 **a** $\frac{2(n-1)\sigma^4}{n^2}$
8.131 $c = \frac{1}{n+1}$
8.133 **b** $\frac{2\sigma^4}{n_1 + n_2 - 2}$

Capítulo 9

9.1 1/3; 2/3; 3/5
9.3 **b** $\frac{12n^2}{(n+2)(n+1)^2}$
9.5 $n - 1$
9.7 $1/n$
9.9 **a** $X_6 = 1$
9.23 **c** necesita $\text{Var}(X_{2i} - X_{2i-1}) < \infty$
9.25 **b** .6826
c No
9.31 $\alpha\beta$
9.35 **a** $\bar{Y}_n$ es insesgada para μ.
b $V(\bar{Y}_n) = \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2$
9.47 $\sum_{i=1}^{n} \ln(Y_i)$; no
9.57 Sí
9.59 $3\left[\bar{Y}^2 + \bar{Y}\left(1 - \frac{1}{n}\right)\right]$
9.61 $\left(\frac{n+1}{n}\right)Y_{(n)}$
9.63 **b** $\frac{3n+1}{3n}Y_{(n)}$
9.69 $\hat{\theta} = \frac{2\bar{Y} - 1}{1 - \bar{Y}}$, no, no es MVUE
9.71 $\hat{\sigma}^2 = m_2' = \frac{1}{n}\sum_{i=1}^n Y_i^2$.
9.75 Con $m_2' = \frac{1}{n}\sum_{i=1}^n Y_i^2$, el estimador MOM de θ es $\hat{\theta} = \frac{1 - 2m_2'}{4m_2' - 1}$.
9.77 $\frac{2}{3}\bar{Y}$
9.81 $\bar{Y}^2$
9.83 **a** $\hat{\theta} = \frac{1}{2}\left(Y_{(n)} - 1\right)$
b $\left(Y_{(n)}\right)^2/12$
9.85 **a** $\hat{\theta} = \frac{1}{\alpha}\bar{Y}$
b $E(\hat{\theta}) = \theta$, $V(\hat{\theta}) = \theta^2/(n\alpha)$
d $\sum_{i=1}^n Y_i$
e $\left(\frac{2\sum_{i=1}^n Y_i}{31.4104}, \frac{2\sum_{i=1}^n Y_i}{10.8508}\right)$
9.87 $\hat{p}_A = .30$, $\hat{p}_B = .38$
$\hat{p}_C = .32$; $-.08 \pm .1641$
9.91 $Y_{(n)}/2$
9.93 **a** $Y_{(1)}$
c $[(\alpha/2)^{1/2n}Y_{(1)}, (1 - (\alpha/2))^{1/2n}Y_{(1)}]$
9.97 **a** $1/\bar{Y}$
b $1/\bar{Y}$

9.99 $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

9.101 $\exp(-\bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\bar{Y}\exp(-2\bar{Y})}{n}}$

9.103 $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$

9.105 $\hat{\sigma}^2 = \frac{\Sigma(Y_i-\mu)^2}{n}$

9.107 $\exp(-t/\bar{Y})$

9.109 **a** $\hat{N}_1 = 2\bar{Y} - 1$
b $\frac{N^2-1}{3n}$

9.111 252 ± 85.193

Capítulo 10

10.3 **a** $c = 11$
b .596
c .057

10.5 $c = 1.684$

10.7 **a** Falso
b Falso
c Verdadero
d Verdadero
e Falso
f **i** Verdadero
ii Verdadero
iii Falso

10.17 **a** H_0: $\mu_1 = \mu_2$, H_a: $\mu_1 > \mu_2$
c $z = .075$

10.21 $z = 3.65$, rechazar H_0

10.23 **a-b** H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 \neq 0$, que es una prueba de dos colas.
c $z = -.954$, que no lleva a un rechazo con $\alpha = .10$.

10.25 $|z| = 1.105$, no rechazar

10.27 $z = -.1202$, no rechazar

10.29 $z = 4.47$

10.33 $z = 1.50$, no

10.35 $z = -1.48$ (1 = sin hogar), no

10.37 aprox. 0 (.0000317)

10.39 .6700

10.41 .025

10.43 **a** .49
b .1056

10.45 $.22 \pm .155$ o (.065, .375)

10.47 .5148

10.49 129.146, sí

10.51 $z = 1.58$ valor $p = .1142$, no rechazar

10.53 **a** $z = -.996$, valor $p = .0618$
b No
c $z = -1.826$, valor $p = .0336$
d Sí

10.55 $z = -1.538$; valor $p = .0616$; no rechazar H_0 con $\alpha = .01$

10.57 $z = -1.732$; valor $p = .0836$

10.63 **a** $t = -1.341$, no rechazar H_0

10.65 **a** $t = -3.24$, valor $p < .005$, sí
b Usando la aplicación, .00241
c 39.556 ± 3.55

10.67 **a** $t = 4.568$ y $t_{.01} = 2.821$ de modo que rechace H_0.
b El límite inferior de confianza de 99% es $358 - 2.821\frac{54}{\sqrt{10}} = 309.83$.

10.69 **a** $t = -1.57$, $.10 <$ valor $p < .20$, no rechazar; usando aplicación, valor $p = .13008$
i $-t_{.10} = -1.319$ y $-t_{.05} = -1.714$; $.10 <$ valor $p < .20$.
ii Usando la aplicación, $2P(T < -1.57) = 2(.06504) = .13008$.

10.71 **a** $\bar{y}_1 = 97.856$, $s_1^2 = .3403$, $\bar{y}_2 = 98.489$, $s_2^2 = .3011$, $t = -2.3724$, $-t_{.01} = -2.583$, $-t_{.025} = -2.12$, así $.02 <$ valor $p < .05$
b Usando la aplicación, .03054

10.73 **a** $t = 1.92$, no rechazar $.05 <$ valor $p < .10$; valor p de aplicación $= .07084$
b $t = .365$, no rechazar valor $p > .20$; valor p de aplicación $= .71936$

10.75 $t = -.647$, no rechazar

10.77 **a** $t = -5.54$, rechazar, valor $p < .01$; valor p aprox. de aplicación
b Sí
c $t = 1.56$, $.10 <$ valor $p < .20$; valor p de aplicación $= .12999$
d Sí

10.79 **a** $\chi^2 = 12.6$, no rechazar
b $.05 <$ valor $p < .10$
c Valor p de aplicación $= .08248$

10.83 **a** $\sigma_1^2 \neq \sigma_2^2$
b $\sigma_1^2 < \sigma_2^2$
c $\sigma_1^2 > \sigma_2^2$

10.85 $\chi^2 = 22.45$, valor $p < .005$; valor p de aplicación $= .0001$

10.89 **a** .15
b .45
c .75
d 1

10.91 **a** Rechazar si $\bar{Y} > 7.82$.
b .2611, .6406, .9131, .9909

10.93 $n = 16$

10.95 **a** $U = \frac{2}{\beta_0} \sum_{i=1}^{4} Y_i$ tiene $\chi^2_{(24)}$ distribución bajo H_0: rechazar H_0 si $U > \chi^2_\alpha$
b Sí

10.97 **d** Sí, es UMP

10.99 **a** $\sum_{i=1}^{n} Y_i \geq k$
b Use tabla de Poisson para hallar k tal que $P(\sum Y_i \geq k) = \alpha$
c Sí

10.101 **a** $\sum_{i=1}^{n} Y_i < c$
b Sí

10.103 **a** Rechazar H_0 si $Y_{(n)} \leq \theta_0 \sqrt[n]{\alpha}$
b Sí

10.107 $\chi^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{\sigma_0^2}$ tiene $\chi^2_{(n+m-2)}$ distribución bajo H_0; rechazar si $\chi^2 > \chi^2_\alpha$

10.109 **a** $\lambda = \dfrac{(\bar{X})^m (\bar{Y})^m}{\left(\dfrac{m\bar{X} + n\bar{Y}}{m+n}\right)^{m+n}}$
b $\bar{X}/\bar{Y}$ distribuida como F con $2m$ y $2n$ grados de libertad

10.115 **a** Verdadero
b Falso
c Falso
d Verdadero
e Falso
f Falso
g Falso
h Falso
i Verdadero

10.117 **a** $t = -22.17$, valor $p < .01$
b $-.0105 \pm .001$
c Sí
d No

10.119 **a** H_0: $p = .20$, H_a: $p > .20$
b $\alpha = .0749$

10.121 $z = 5.24$, valor p aprox. 0

10.123 **a** $F = 2.904$, no
b (.050, .254)

10.125 **a** $t = -2.657$, $.02 <$ valor $p < .05$
b -4.542 ± 3.046

10.127 $T = \dfrac{(\bar{X}+\bar{Y}-\bar{W})-(\mu_1-\mu_2-\mu_3)}{\left\{\left(\frac{1+a+b}{n(3n-3)}\right)\left[\sum(X_i-\bar{X})^2 + \frac{1}{a}\sum(Y_i-\bar{Y})^2 + \frac{1}{b}\sum(W_i-\bar{W})^2\right]\right\}^{1/2}}$ con $(3n-3)$ grados de libertad

10.129 $\lambda = \left(\dfrac{\sum_{i=1}^{n}(y_i - y_{(1)})}{n\theta_{1,0}}\right)^n \times \exp\left[-\dfrac{\sum_{i=1}^{n}(y_i - y_{(1)})}{\theta_{1,0}} + n\right]$.

Capítulo 11

11.3 $\hat{y} = 1.5 - .6x$

11.5 $\hat{y} = 21.575 + 4.842x$

11.7 **a** La relación parece ser proporcional a x^2.
b No
c No, es el mejor modelo *lineal*.

11.9 **b** $\hat{y} = -15.45 + 65.17x$
d 108.373

11.11 $\hat{\beta}_1 = 2.514$

11.13 **a** La recta de mínimos cuadrados es $\hat{y} = 452.119 - 29.402x$

11.17 **a** SSE = 18.286; $S^2 = 18.286/6 = 3.048$
b La recta ajustada es $\hat{y} = 43.35 + 2.42x^*$. Se encuentra la misma respuesta para SSE (y por tanto S^2).

11.19 **a** La recta de mínimos cuadrados es: $\hat{y} = 3.00 + 4.75x$
c $s^2 = 5.025$

11.23 **a** $t = -5.20$, rechazar H_0
b $.01 <$ valor $p < .02$
c .01382
d $(-.967, -.233)$

11.25 **a** $t = 3.791$, valor $p < .01$
b valor p de aplicación $= .0053$
c Rechazar
d $.475 \pm .289$

11.29 $T = \dfrac{\hat{\beta}_1 - \hat{\gamma}_1}{S\sqrt{\left(\frac{1}{S_{xx}} + \frac{1}{S_{cc}}\right)}}$, donde $S =$ $(\text{SSE}_Y + \text{SSE}_W)/(n + m - 4)$. H_0 es rechazada a favor de H_a para valores grandes de $|T|$.

11.31 $t = 73.04$, valor p aprox. 0, H_0 es rechazada

11.33 $t = 9.62$, sí

11.35 $x^* = \bar{x}$.

11.37 (4.67, 9.63)

11.39 25.395 ± 2.875

11.41 **b** (72.39, 75.77)

11.43 (59.73, 70.57)

11.45 (−.86, 15.16)

11.47 (.27, .51)

11.51 $t = 9.608$, valor $p < .01$

11.53 **a** $r^2 = .682$
b .682
c $t = 4.146$, rechazar
d Valor p de aplicación = .00161

11.57 **a** signo para r
b r y n

11.59 $r = -.3783$

11.61 $.979 \pm .104$

11.63 **a** $\hat{\beta}_1 = -.0095, \hat{\beta}_0 = 3.603$ y $\hat{\alpha}_1 = -(-.0095) = .0095$, $\hat{\alpha}_0 = \exp(3.603) = 36.70$. Por lo tanto, la ecuación de predicción es $\hat{y} = 36.70e^{-.0095x}$.
b El CI de 90% para α_0 es $\left(e^{3.5883}, e^{3.6171}\right) = (36.17, 37.23)$

11.67 $\hat{y} = 2.1 - .6x$

11.69 **a** $\hat{y} = 32.725 + 1.812x$
b $\hat{y} = 35.5625 + 1.8119x - .1351x^2$

11.73 $t = 1.31$, no rechazar

11.75 21.9375 ± 3.01

11.77 Siguiendo el Ej. 11.76, el PI de 95% es 39.9812 ± 213.807

11.79 21.9375 ± 6.17

11.83 **a** $F = 21.677$, rechazar
b $\text{SSE}_R = 1908.08$

11.85 **a** $F = 40.603$, valor $p < .005$
b 950.1676

11.87 **a** $F = 4.5$, $F_1 = 9.24$, no rechazar H_0
c $F = 2.353$, $F_1 = 2.23$, rechazar H_0

11.89 **a** Verdadero
b Falso
c Falso

11.91 $F = 10.21$

11.93 90.38 ± 8.42

11.95 **a** $\hat{y} = -13.54 - 0.053x$
b $t = -6.86$
c $.929 \pm .33$

11.97 **a** $\hat{y} = 1.4825 + .5x_1 + .1190x_2 - .5x_3$
b $\hat{y} = 2.0715$
c $t = -13.7$, rechazar
d (1.88, 2.26)
e (1.73, 2.41)

11.99 Si $-9 \leq x \leq 9$, escoja $n/2$ en $x = -9$ y $n/2$ en $x = 9$.

11.101 **a** $\hat{y} = 9.34 + 2.46x_1 + .6x_2 + .41x_1x_2$
b 9.34 , 11.80
d Para bacterias A, $\hat{y} = 9.34$. Para bacterias B, $\hat{y} = 11.80$. Los crecimientos observados fueron 9.1 y 12.2, respectivamente.
e $12.81 \pm .37$
f $12.81 \pm .78$

11.107 **a** $r = .89$
b $t = 4.78$, valor $p < .01$, rechazar

Capítulo 12

12.1 $n_1 = 34, n_2 = 56$

12.3 $n = 246, n_1 = 93, n_2 = 154$

12.5 Con $n = 6$, tres ratas deberían recibir $x = 2$ unidades y tres ratas deberían recibir $x = 5$ unidades.

12.11 **a** Esto ocurre cuando $\rho > 0$.
b Esto ocurre cuando $\rho = 0$.
c Esto ocurre cuando $\rho < 0$.
d Pareados mejor cuando $\rho > 0$, independiente mejor cuando $\rho < 0$

12.15 **a** $t = 2.65$, rechazar

12.17 **a** μ_i

12.31 **a** μ_i
b $\mu_i, \dfrac{1}{n}[\sigma_P^2 + \sigma^2]$
c $\mu_1 - \mu_2, 2\sigma^2/n$, normal

12.35 **a** $t = -4.326$, $.01 <$ valor $p < .025$
b -1.58 ± 1.014
c 65 pares

12.37 $k_1 = k_3 = .25; k_2 = .50$

Capítulo 13

13.1 **a** $F = 2.93$, no rechazar
b .109
c $|t| = 1.71$, no rechazar, $F = t^2$

13.7 **a** $F = 5.2002$, rechazar
b valor $p = .01068$

13.9 SSE = .020; $F = 2.0$, no rechazar

13.11 SST = .7588; SSE = .7462; $F = 19.83$, valor $p < .005$, rechazar

13.13 SST = 36.286; SSE = 76.6996; $F = 38.316$, valor $p < .005$, rechazar

13.15 $F = 63.66$, sí, valor $p < .005$

13.21 **a** -12.08 ± 10.96
b Más largo
c Menos grados de libertad

13.23 **a** $1.568 \pm .164$ o $(1.404, 1.732)$; sí
b $(-.579, -.117)$; sí

13.25 $.28 \pm .102$

13.27 **a** 95% de CI para μ_A: 76 ± 8.142 o $(67.868, 84.142)$
b 95% de CI para μ_B: 66.33 ± 10.51 o $(55.82, 76.84)$
c 95% de CI para $\mu_A - \mu_B$: 9.667 ± 13.295

13.29 **a** $6.24 \pm .318$
b $-.29 \pm .241$

13.31 **a** $F = 1.32$, no
b $(-.21, 4.21)$

13.33 $(1.39, 1.93)$

13.35 **a** 2.7 ± 3.750
b 27.5 ± 2.652

13.37 **a** μ
b Media general

13.39 **b** $(2\sigma^2)/b$

13.41 **a** $F = 3.11$, no rechazar
b valor $p > .10$
c valor $p = .1381$
d $s_D^2 = 2\text{MSE}$

13.45 **a** $F = 10.05$; rechazar
b $F = 10.88$; rechazar

13.47

Fuente	gl	SS	MS	F
Tratamientos	3	8.1875	2.729	1.40
Bloques	3	7.1875	2.396	1.23
Error	9	17.5625	1.95139	
Total	15	32.9375		

$F = 1.40$, no rechazar

13.49 $F = 6.36$; rechazar

13.53 El CI de 95% es 2 ± 2.83.

13.55 El CI de 95% es $.145 \pm .179$.

13.57 El CI de 95% es -4.8 ± 5.259.

13.59 $n_A \geq 3$

13.61 $b = 16$; $n = 48$

13.63 Los tamaños muestrales difieren.

13.69 **a** $\beta_0 + \beta_3$ es la respuesta media al tratamiento A en el bloque III
b β_3 es la diferencia en respuestas medias a los químicos A y D en el bloque III.

13.71 $F = 7$; H_0 es rechazada

13.73 Tan homogénea como sea posible dentro de bloques

13.75 **b** $F = 1.05$; no rechazar

13.77 **a** A 95% CI es $.084 \pm .06$ o $(.024, .144)$.

13.79 **a** 16
b 135 grados de libertad dejados para error.
c 14.14

13.81 $F = 7.33$; sí; el bloqueo induce pérdida en grados de libertad para estimar σ^2; podría resultar en pérdida de vista de información si la variación de bloque a bloque es pequeña

13.83 **a**

Fuente	gl	SS	MS	F
Tratamientos	2	524,177.167	262,088.58	258.237
Bloques	3	173,415	57,805.00	56.95
Error	6	6,089.5	1,014.9167	
Total	11	703,681.667		

b 6
c Sí, $F = 258.19$, valor $p < .005$
d Sí, $F = 56.95$, valor $p < .005$
e 22.527
f -237.25 ± 55.13

13.85 **a** SST = 1.212, gl = 4
SSE = .571, gl = 22
$F = 11.68$; valor $p < .005$
b $|t| = 2.73$; H_0 es rechazada; $2(.005) <$ valor $p < 2(.01)$.

13.87 Cada intervalo debe tener coeficiente de confianza $1 - .05/4 = .9875 \approx .99$;
$\mu_A - \mu_D$: $.320 \pm .251$
$\mu_B - \mu_D$: $.145 \pm .251$
$\mu_C - \mu_D$: $.023 \pm .251$
$\mu_E - \mu_D$: $-.124 \pm .251$

13.89 **b** σ_β^2
c $\sigma_\beta^2 = 0$
13.91 **a** μ; $\sigma_\beta^2 + \frac{1}{k}\sigma_\varepsilon^2$
b $\sigma_\beta^2 + \left(\frac{b}{k-1}\right)\sum_{i=1}^{k} \tau_i^2$
c $\sigma_\varepsilon^2 + k\sigma_B^2$
d σ_ε^2

Capítulo 14

14.1 **a** $X^2 = 3.696$, no rechazar
b Valor p de aplicación $= .29622$
14.3 $X^2 = 24.48$, valor $p < .005$
14.5 **a** $z = 1.50$, no rechazar
b Hipótesis sugerida por datos observados
14.7 $.102 \pm .043$
14.9 **a** $.39 \pm .149$
b $.37 \pm .187$, $.39 \pm .182$, $.48 \pm .153$
14.11 $X^2 = 69.42$, rechazar
14.13 **a** $X^2 = 18.711$, rechazar
b valor $p < .005$
c valor p de aplicación $= .00090$
14.15 **b** X^2 también multiplicada por k
14.17 **a** $X^2 = 19.0434$ con valor p de .004091.
b $X^2 = 60.139$ con valor p de aproximadamente 0.
c Algunas cuentas esperadas < 5
14.19 **a** $X^2 = 22.8705$, rechazar
b valor $p < .005$
14.21 **a** $X^2 = 13.99$, rechazar
b $X^2 = 13.99$, rechazar
c $X^2 = 1.36$, no rechazar
14.25 **b** $X^2 = 19.1723$, valor $p =$ 0.003882, rechazar
c $-.11 \pm .135$
14.27 $X^2 = 38.43$, sí
14.29 **a** $X^2 = 14.19$, rechazar
14.31 $X^2 = 21.51$, rechazar
14.33 $X^2 = 6.18$, rechazar; $.025 <$ valor $p < .05$
14.35 **a** Sí
b valor $p = .002263$
14.37 $X^2 = 8.56$, gl $= 3$; rechazar
14.41 $X^2 = 3.26$, no rechazar
14.43 $X^2 = 74.85$, rechazar

Capítulo 15

15.1

Región de rechazo	α
$M \le 6$ o $M \ge 19$	$P(M \le 6) + P(M \ge 19) = .014$
$M \le 7$ o $M \ge 18$	$P(M \le 7) + P(M \ge 18) = .044$
$M \le 8$ o $M \ge 17$	$P(M \le 8) + P(M \ge 17) = .108$

15.3 **a** $m = 2$, sí
b Varianzas no iguales
15.5 $P(M \le 2 \text{ o } M \ge 8) = .11$, no
15.7 **a** $P(M \le 2 \text{ o } M \ge 7) = .18$, no rechazar
b $t = -1.65$, no rechazar
15.9 **a** valor $p = .011$, no rechazar
15.11 $T = \text{mín}(T^+, T^-)$, $T = T^-$.
15.13 **a** $T = 6$, $.02 <$ valor $p < .05$
b $T = 6$, $0.1 <$ valor $p < .025$
15.15 $T = 3.5$, $.025 <$ valor $p < .05$
15.17 $T = 11$, rechazar
15.21 **a** $U = 4$; valor $p = .0364$
b $U = 35$; valor $p = .0559$
c $U = 1$; valor $p = .0476$
15.23 $U = 9$, no rechazar
15.25 $z = -1.80$, rechazar
15.27 $U = 0$, valor $p = .0096$
15.29 $H = 16.974$, valor $p <. 001$
15.31 **a** SST = 2586.1333; SSE = 11,702.9; $F = 1.33$, no rechazar
b $H = 1.22$, no rechazar
15.33 $H = 2.03$, no rechazar
15.37 **a** No, valor $p = .6685$
b No rechazar H_0
15.39 $F_r = 6.35$, rechazar
15.41 **a** $F_r = 65.675$, valor $p < .005$, rechazar
b $m = 0$, $P(M = 0) = 1/256$, valor $p = 1/128$
15.45 La distribución nula está dada por $P(F_r = 0) = P(F_r = 4) = 1/6$ y $P(F_r = 1) = P(F_r = 3) = 1/3$.
15.47 $R = 6$, no

15.49 **a** .0256
b Un número normalmente pequeño de lotes (juzgados a $\alpha = .05$) implicaría una agrupación de piezas defectuosas en tiempo; no rechazar.
15.51 $R = 13$, no rechazar
15.53 $r_S = .911818$; sí.
15.55 **a** $r_S = -.8449887$
b Rechazar
15.57 $r_S = .6768$, usar prueba de dos colas, rechazar
15.59 $r_S = 0$; valor $p < .005$
15.61 **a** Diseño aleatorizado de bloque
b No
c valor $p = .04076$, sí
15.63 $T = 73.5$, no rechazar, consistente con el Ej. 15.62
15.65 $U = 17.5$, no rechazar H_0
15.67 .0159
15.69 $H = 7.154$, rechazar
15.71 $F_r = 6.21$, no rechazar
15.73 .10

Capítulo 16

16.1 **a** $\beta(10, 30)$
b $n = 25$
c $\beta(10, 30)$, $n = 25$
d Sí
e Posterior para $\beta(1, 3)$ previa.
16.3 **c** Las medias se acercan a .4, desv. est. Disminuye
e Se ve cada vez más como distribución normal.
16.7 **a** $\dfrac{Y+1}{n+4}$
b $\dfrac{np+1}{n+4}$; $\dfrac{np(1-p)}{(n+4)^2}$
16.9 **b** $\dfrac{\alpha+1}{\alpha+\beta+Y}$; $\dfrac{(\alpha+1)(\beta+Y-1)}{(\alpha+\beta+Y+1)(\alpha+\beta+Y)}$
16.11 **e** $\bar{Y}\left(\dfrac{n\beta}{n\beta+1}\right) + \alpha\beta\left(\dfrac{1}{n\beta+1}\right)$
16.13 **a** (.099, .710)
b Ambas probabilidades son .025
c $P(.099 < p < .710) = .95$
h Más corta para n más grande.
16.15 (.06064, .32665)
16.17 (.38475, .66183)
16.19 (5.95889, 8.01066)
16.21 Probabilidades posteriores de nulo y alternativa son .9526 y .0474, respectivamente, aceptar H_0.
16.23 Probabilidades posteriores de nulo y alternativa son .1275 y .8725, respectivamente, aceptar H_a.
16.25 Probabilidades posteriores de nulo y alternativa son .9700 y .0300, respectivamente, aceptar H_0.

ÍNDICE

E

F

J

L

M

N

O

P

U

V

W

Z

Distribuciones continuas

Distribución	Función de probabilidad	Media	Varianza	Función de generación de momento
Uniforme	$f(y) = \frac{1}{\theta_2 - \theta_1}; \theta_1 \le y \le \theta_2$	$\frac{\theta_1 + \theta_2}{2}$	$\frac{(\theta_2 - \theta_1)^2}{12}$	$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$ $-\infty < y < +\infty$	μ	σ^2	$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$
Exponencial	$f(y) = \frac{1}{\beta} e^{-y/\beta}; \quad \beta > 0$ $0 < y < \infty$	β	β^2	$(1 - \beta t)^{-1}$
Gamma	$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right] y^{\alpha-1} e^{-y/\beta};$ $0 < y < \infty$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha}$
Ji cuadrada	$f(y) = \frac{(y)^{(v/2)-1} e^{-y/2}}{2^{v/2}\Gamma(v/2)};$ $y^2 > 0$	v	$2v$	$(1-2t)^{-v/2}$
Beta	$f(y) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] y^{\alpha-1}(1-y)^{\beta-1};$ $0 < y < 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	no existe en la forma cerrada

Distribuciones discretas

Distribución	Función de probabilidad	Media	Varianza	Función de generación de momento
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$; $y = 0, 1, \ldots, n$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geométrica	$p(y) = p(1-p)^{y-1}$; $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hipergeométrica	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$; $y = 0, 1, \ldots, n$ if $n \leq r$, $y = 0, 1, \ldots, r$ if $n > r$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$; $y = 0, 1, 2, \ldots$	λ	λ	$\exp[\lambda(e^t - 1)]$
Binomial negativa	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$; $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$

En este exitoso texto, el más ampliamente utilizado en las 200 mejores escuelas de Estados Unidos, los autores Dennis D. Wackerly, William Mendenhall III y Richard L. Scheaffer, presentan la teoría de la estadística en el contexto de la solución de problemas prácticos y aplicaciones al mundo real. Utilizan eficientemente las matemáticas como una herramienta necesaria para promover una firme comprensión de las técnicas estadísticas. Este planteamiento práctico le ayuda al lector a descubrir la naturaleza de la estadística y a comprender rápidamente su importante papel en la investigación científica.

¿Cómo aprovechar al máximo este texto?

- Los autores *enfatizan la conectividad*, explican cómo los grandes temas juegan un importante papel en la inferencia estadística y cómo se relacionan entre sí, asegurando la lectura de las introducciones y los resúmenes de capítulo.
- *Los conceptos importantes se agrupan como definiciones.* Léalos y vuélvalos a leer hasta que los entienda, así como el marco en el que todos ellos se desarrollan.
- *Trabaje los ejercicios para poner su nuevo conocimiento a prueba.* Ya que muchos ejercicios están basados en datos reales o en escenarios experimentales actuales, esto le permite ver los diferentes usos prácticos de varios métodos estadísticos y probabilísticos.
- *Utilice la colección de* ***applets*** *interactivos en línea* (simulaciones) para desarrollar su conocimiento y la comprensión de los conceptos clave.

Por todo esto y mucho más que encontrará al interior del libro, estamos seguros que lo convertirá en su libro de cabecera para sus cursos de estadística.

CENGAGE Learning™

http://latinoamerica.cengage.com

ISBN-13: 978-607481399-9
ISBN-10: 607481399-X

9 786074 813999