CENGAGE Learning™

PROBABILIDAD Y ESTADÍSTICA

para ingeniería y ciencias

Octava edición

JAY L. DEVORE

Probabilidad y estadística para ingeniería y ciencias

OCTAVA EDICIÓN

Probabilidad y estadística para ingeniería y ciencias

JAY DEVORE

California Polytechnic State University, San Luis Obispo

Traducción:
Patricia Solorio Gómez
Traductora profesional

Revisión técnica:
Ana Elizabeth García Hernández
Universidad LaSalle Morelia

CENGAGE Learning™

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

CENGAGE
Learning™

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

Presidente de Cengage Learning Latinoamérica:
Fernando Valenzuela Migoya

Director de producto y desarrollo Latinoamérica:
Daniel Oti Yvonnet

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

Editor:
Sergio R. Cervantes González

Coordinadora de producción editorial:
Abril Vega Orozco

Editor de producción:
Timoteo Eliosa García

Coordinador de manufactura:
Rafael Pérez González

Diseño de portada:
Rokusek Design

Composición tipográfica:
Imagen Editorial

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

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

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

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

Para mi nieto
Philip, quien es
estadísticamente
significativo.

Contenido

1 Generalidades y estadística descriptiva

Introducción 1
1.1 Poblaciones, muestras y procesos 2
1.2 Métodos pictóricos y tabulares en la estadística descriptiva 12
1.3 Medidas de ubicación 28
1.4 Medidas de variabilidad 35
Ejercicios suplementarios 46
Bibliografía 49

2 Probabilidad

Introducción 50
2.1 Espacios muestrales y eventos 51
2.2 Axiomas, interpretaciones, y propiedades de la probabilidad 55
2.3 Técnicas de conteo 64
2.4 Probabilidad condicional 73
2.5 Independencia 83
Ejercicios suplementarios 88
Bibliografía 91

3 Variables aleatorias discretas y distribuciones de probabilidad

Introducción 92
3.1 Variables aleatorias 93
3.2 Distribuciones de probabilidad para variables aleatorias discretas 96
3.3 Valores esperados 106
3.4 Distribución de probabilidad binomial 114
3.5 Distribuciones hipergeométrica y binomial negativa 122
3.6 Distribución de probabilidad de Poisson 128
Ejercicios suplementarios 133
Bibliografía 136

4 Variables aleatorias continuas y distribuciones de probabilidad

Introducción 137
4.1 Funciones de densidad de probabilidad 138
4.2 Funciones de distribución acumulativa y valores esperados 143
4.3 Distribución normal 152
4.4 Distribuciones exponencial y gamma 165
4.5 Otras distribuciones continuas 171
4.6 Gráficas de probabilidad 178
Ejercicios suplementarios 188
Bibliografía 192

5 Distribuciones de probabilidad conjunta y muestras aleatorias

Introducción 193
5.1 Variables aleatorias conjuntamente distribuidas 194
5.2 Valores esperados, covarianza y correlación 206
5.3 Estadísticos y sus distribuciones 212
5.4 Distribución de la media muestral 223
5.5 Distribución de una combinación lineal 230
Ejercicios suplementarios 235
Bibliografía 238

6 Estimación puntual

Introducción 239
6.1 Algunos conceptos generales de estimación puntual 240
6.2 Métodos de estimación puntual 255
Ejercicios suplementarios 265
Bibliografía 266

7 Intervalos estadísticos basados en una sola muestra

Introducción 267
7.1 Propiedades básicas de los intervalos de confianza 268
7.2 Intervalos de confianza de muestra grande para una media y proporción de población 276

7.3 Intervalos basados en una distribución de población normal 285
7.4 Intervalos de confianza para la varianza y desviación estándar de una población normal 294
Ejercicios suplementarios 297
Bibliografía 299

8 Pruebas de hipótesis basadas en una sola muestra

Introducción 300
8.1 Hipótesis y procedimientos de prueba 301
8.2 Pruebas sobre una media de población 310
8.3 Pruebas relacionadas con una proporción de población 323
8.4 Valores *P* 328
8.5 Algunos comentarios sobre la selección de una prueba 339
Ejercicios suplementarios 342
Bibliografía 344

9 Inferencias basadas en dos muestras

Introducción 345
9.1 Pruebas *z* e intervalos de confianza para una diferencia entre dos medias de población 346
9.2 Prueba *t* con dos muestras e intervalo de confianza 357
9.3 Análisis de datos pareados 365
9.4 Inferencias sobre una diferencia entre proporciones de población 375
9.5 Inferencias sobre dos varianzas de población 382
Ejercicios suplementarios 386
Bibliografía 390

10 Análisis de la varianza

Introducción 391
10.1 ANOVA unifactorial 392
10.2 Comparaciones múltiples en ANOVA 402
10.3 Más sobre ANOVA unifactorial 408
Ejercicios suplementarios 417
Bibliografía 418

11 Análisis multifactorial de la varianza

Introducción 419

11.1 ANOVA bifactorial con $K_{ij} = 1$ 420

11.2 ANOVA bifactorial con $K_{ij} > 1$ 433

11.3 ANOVA con tres factores 442

11.4 Experimentos 2^p factoriales 451

Ejercicios suplementarios 464

Bibliografía 467

12 Regresión lineal simple y correlación

Introducción 468

12.1 Modelo de regresión lineal simple 469

12.2 Estimación de parámetros de modelo 477

12.3 Inferencias sobre el parámetro de pendiente β_1 490

12.4 Inferencias sobre $\mu_{Y \cdot x^*}$ y predicción de valores Y futuros 499

12.5 Correlación 508

Ejercicios suplementarios 518

Bibliografía 522

13 Regresión múltiple y no lineal

Introducción 523

13.1 Aptitud y verificación del modelo 524

13.2 Regresión con variables transformadas 531

13.3 Regresión polinomial 543

13.4 Análisis de regresión múltiple 553

13.5 Otros problemas en regresión múltiple 574

Ejercicios suplementarios 588

Bibliografía 593

14 Pruebas de bondad de ajuste y análisis de datos categóricos

Introducción 594

14.1 Pruebas de bondad de ajuste cuando las probabilidades categóricas se satisfacen por completo 595

14.2 Pruebas de bondad de ajuste para hipótesis compuestas 602
14.3 Tablas de contingencia mutuas (o bidireccionales) 613
Ejercicios suplementarios 621
Bibliografía 624

15 Procedimientos de distribución libre

Introducción 625
15.1 La prueba Wilcoxon de rango con signo 626
15.2 Prueba Wilcoxon de suma de rangos 634
15.3 Intervalos de confianza de distribución libre 640
15.4 ANOVA de distribución libre 645
Ejercicios suplementarios 649
Bibliografía 650

16 Métodos de control de calidad

Introducción 651
16.1 Comentarios generales sobre gráficas de control 652
16.2 Gráficas de control para ubicación de proceso 654
16.3 Gráficas de control para variación de proceso 663
16.4 Gráficas de control para atributos 668
16.5 Procedimientos CUSUM 672
16.6 Muestreo de aceptación 680
Ejercicios suplementarios 686
Bibliografía 687

Tablas de apéndice

A.1 Distribución binomial acumulada A-2
A.2 Distribución acumulada de Poisson A-4
A.3 Áreas de la curva normal estándar A-6
A.4 La función gamma incompleta A-8
A.5 Valores críticos para distribuciones *t* A-9
A.6 Valores críticos de tolerancia para distribuciones normales de población A-10
A.7 Valores críticos para distribuciones chi-cuadrada A-11
A.8 Áreas de cola de la curva *t* A-12
A.9 Valores críticos de la distribución *F* A-14

A.10 Valores críticos para la distribución de rango estudentizado A-20
A.11 Áreas de cola de la curva chi-cuadrada A-21
A.12 Valores críticos para la prueba de normalidad Ryan-Joiner A-23
A.13 Valores críticos para la prueba Wilcoxon de rangos con signo A-24
A.14 Valores críticos para la prueba Wilcoxon de suma de rangos A-25
A.15 Valores críticos para el intervalo Wilcoxon de rangos con signo A-26
A.16 Valores críticos para el intervalo Wilcoxon de suma de rangos A-27
A.17 Curvas β para pruebas t A-28

Respuestas a ejercicios seleccionados de número impar A-29
Glosario de símbolos y abreviaturas G-1
Índice I-1

Prefacio

Propósito

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

Enfoque

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

Nivel matemático

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

Contenido

El capítulo 1 se inicia con algunos conceptos y terminología básicos (población, muestra, estadística descriptiva e inferencial, estudios enumerativos contra analíticos, y así sucesivamente) y continúa con el estudio de métodos descriptivos gráficos y numéricos importantes. En el capítulo 2 se da un desarrollo un tanto tradicional de la probabilidad, seguido por distribuciones de probabilidad de variables aleatorias continuas y discretas en los capítulos 3 y 4, respectivamente. Las distribuciones conjuntas y sus propiedades se analizan en la primera parte del capítulo 5. La última parte de este capítulo introduce la estadística y sus distribuciones muestrales, las cuales constituyen el puente entre probabilidad e inferencia. Los siguientes tres capítulos se ocupan de la estimación puntual, los intervalos estadísticos y la comprobación de hipótesis basados en una muestra única. Los métodos de inferencia que implican dos muestras independientes y datos apareados se presentan en el capítulo 9. El análisis de la varianza es el tema de los capítulos 10 y 11 (unifactorial y multifactorial, respectivamente). La regresión aparece por primera vez en el capítulo 12 (el modelo de regresión lineal simple y correlación) y regresa para una amplia repetición en el capítulo 13. Los últimos tres capítulos analizan métodos ji al cuadrado, procedimientos sin distribución (no paramétricos) y técnicas de control estadístico de calidad.

Ayuda para el aprendizaje de los estudiantes

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

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

Nuevo en esta edición

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

Reconocimentos

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

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

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

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

Jay Devore

1 Generalidades y estadística descriptiva

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

Max Levchin, cofundador de Paypal, fundador de Slide.

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

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

Hal Varian, economista en jefe Google.

The New York Times, 6 de agosto de 2009.

INTRODUCCIÓN

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

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

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

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

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

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

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

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

1.1 Poblaciones, muestras y procesos

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

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

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

$x =$ marca de la calculadora de un estudiante

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

$z =$ distancia de frenado de un automóvil en condiciones específicas

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

M A A A M A A M A A

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

5.6 5.1 6.2 6.0 5.8 6.5 5.8 5.5

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

Ramas de la estadística

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

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

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

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

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

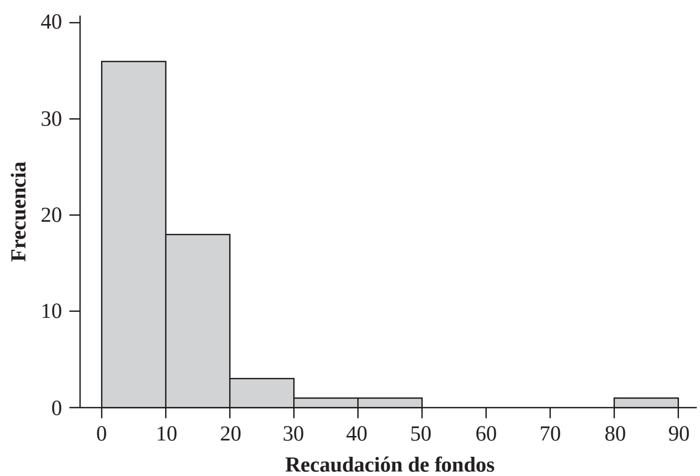


Figura 1.1 Gráfica de tallos y hojas (truncada a diez dígitos) de Minitab e histograma para los datos del porcentaje de recaudación de fondos de caridad

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

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

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

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

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

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

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

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

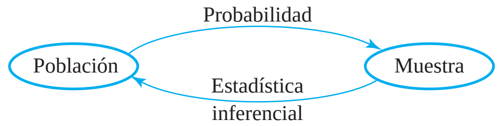


Figura 1.2 Relación entre probabilidad y estadística inferencial

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

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

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

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

El ámbito de la estadística moderna

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

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

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

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

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

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

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

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

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

Estudios enumerativos contra analíticos

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

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

Recopilación de datos

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

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

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

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

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

Ejemplo 1.4 Un artículo en el *New York Times* (27 de enero de 1987) reportó que el riesgo de sufrir un ataque cardiaco podría ser reducido tomando aspirina. Esta conclusión se basó en un experimento diseñado que incluía tanto un grupo de control de individuos que tomaron un placebo que tenía la apariencia de aspirina pero que se sabía era inerte y un grupo de tratamiento que tomó aspirina de acuerdo con un régimen específico. Los sujetos fueron asignados al azar a los grupos para protegerlos contra cualquier prejuicio de modo que se pudieran utilizar métodos basados en la probabilidad para analizar los datos. De los 11,034 individuos en el grupo de control, 189 experimentaron subsecuentemente ataques cardiacos, mientras que sólo 104 de los 11,037 en el grupo de aspirina sufrieron un ataque cardiaco. La tasa de incidencia de ataques cardiacos en el grupo de tratamiento fue de sólo aproximadamente la mitad de aquella en el grupo de control. Una posible explicación de este resultado es la variación de la probabilidad, que la aspirina en realidad no tiene el efecto deseado y la diferencia observada es sólo una variación típica del mismo modo que el lanzamiento al aire de dos monedas idénticas por lo general produciría diferentes cantidades de águilas. No obstante, en este caso, los métodos inferenciales sugieren que la variación de la probabilidad por sí misma no puede explicar en forma adecuada la magnitud de la diferencia observada. ■

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

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

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

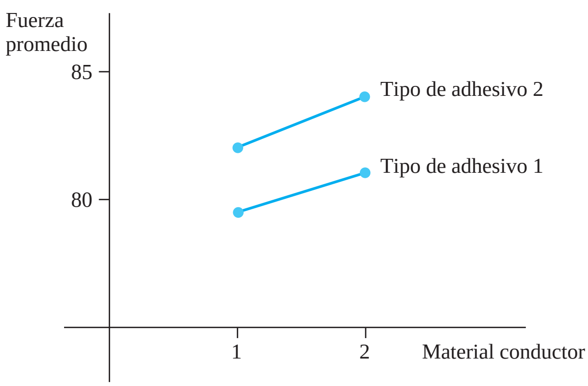


Figura 1.3 Fuerzas adhesivas promedio en el ejemplo 1.5

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

EJERCICIOS Sección 1.1 (1–9)

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

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

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

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

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

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

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

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

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

1.2 Métodos pictóricos y tabulares en la estadística descriptiva

La estadística descriptiva se divide en dos temas generales. En esta sección se considera la representación de un conjunto de datos por medio de técnicas visuales. En las secciones 1.3 y 1.4 se desarrollarán algunas medidas numéricas para conjuntos de datos. Es posible que usted ya conozca muchas técnicas visuales; tablas de frecuencia, hojas de contabilidad, histogramas, gráficas de pastel, gráficas de barras, diagramas de puntos y similares. Aquí se seleccionan algunas de estas técnicas que son más útiles y pertinentes para la probabilidad y estadística inferencial.

Notación

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

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

Gráficas de tallos y hojas

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

Pasos para construir una gráfica de tallos y hojas

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

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

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

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

0	4
1	1345678889
2	1223456666777889999
3	0112233344555666677777888899999
4	111222223344445566666677788888999
5	00111222233455666667777888899
6	01111244455666778

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

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

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

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

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

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

3 | 371 . . .

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

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

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

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

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

64	35	64	33	70			
65	26	27	06	83			
66	05	94	14				
67	90	70	00	98	70	45	13
68	90	70	73	50			
69	00	27	36	04			
70	51	05	11	40	50	22	
71	31	69	68	05	13	65	
72	80	09					

Tallo: dígitos de millares y centenas
Hojas: dígitos de decenas y unidades

(a)

```
          Tallos y hojas de yardaje N = 40
Unidad de hoja = 10
     4        64  3367
     8        65  0228
    11        66  019
    18        67  0147799
   (4)        68  5779
    18        69  0023
    14        70  012455
     8        71  013666
     2        72  08
```

(b)

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

Gráficas de puntos

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

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

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

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

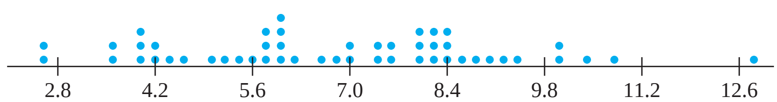


Figura 1.6 Gráfica de puntos para los datos del ejemplo 1.8

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

Histogramas

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

DEFINICIÓN

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

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

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

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

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

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

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

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

Construcción de un histograma para datos discretos

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

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

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

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

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

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

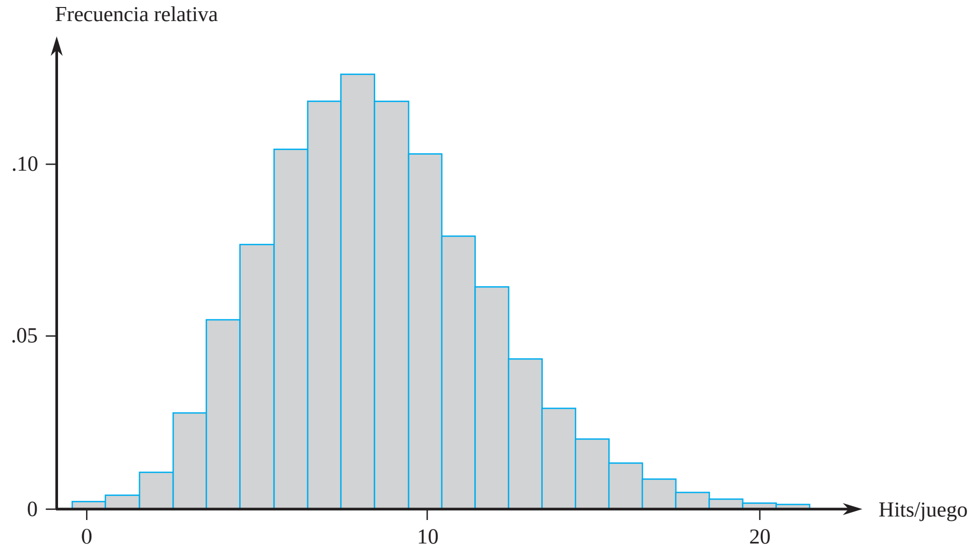


Figura 1.7 Histograma del número de hits por juego de nueve entradas

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

$$\begin{array}{l}\text{proporción de juegos} \\ \text{de dos hits a lo sumo}\end{array} = \begin{array}{c}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 0\end{array} + \begin{array}{c}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 1\end{array} + \begin{array}{c}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 2\end{array}$$

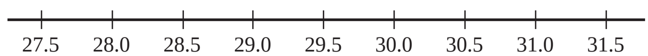
$$= .0010 + .0037 + .0108 = .0155$$

Asimismo,

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

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

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



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

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

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

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

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

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

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

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

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

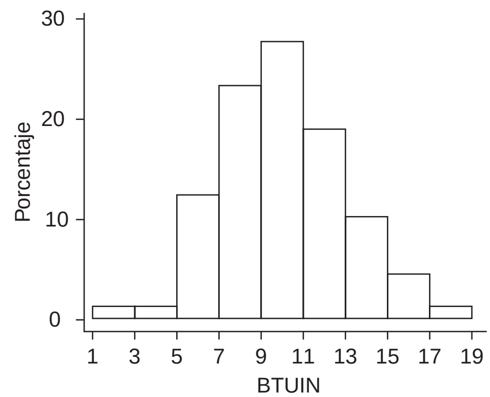


Figura 1.8 Histograma de los datos de consumo de energía del ejemplo 1.10

De acuerdo con el histograma,

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

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

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

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

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

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

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

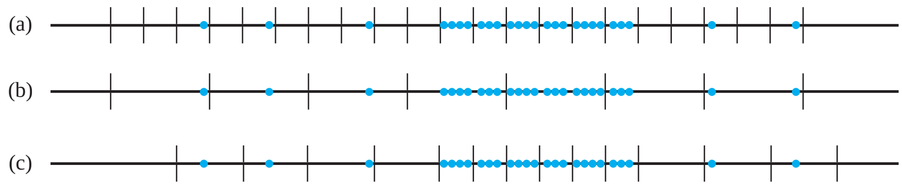


Figura 1.9 Selección de intervalos de clase para datos de "densidad variable": (a) intervalos de ancho igual muy cortos; (b) unos cuantos intervalos de ancho igual; (c) intervalos de ancho desigual

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

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

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

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

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

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

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

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

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

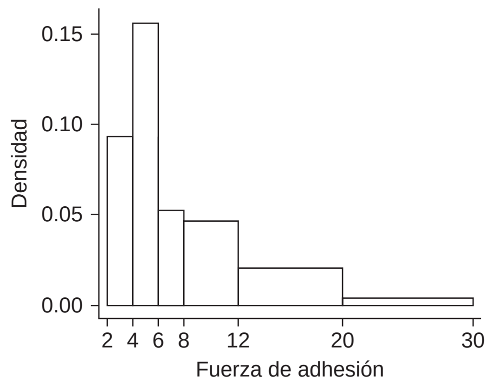


Figura 1.10 Histograma Minitab de densidad para la fuerza de adhesión del ejemplo 1.11 ■

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

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

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

Formas de histograma

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

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

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

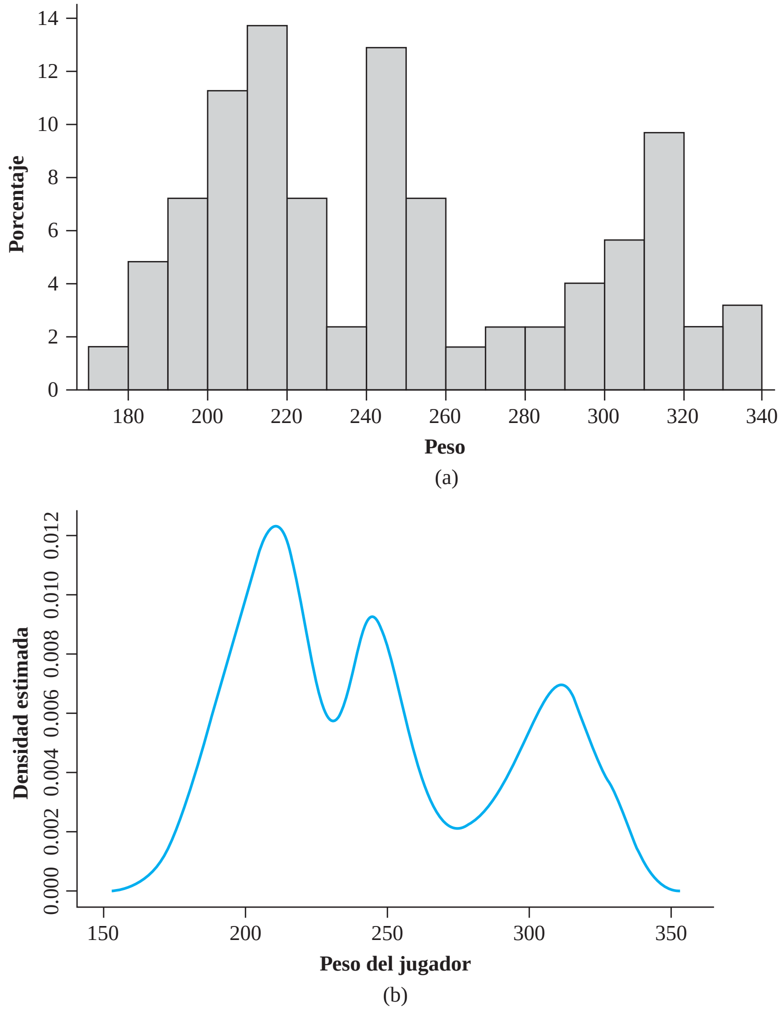


Figura 1.11 Peso de los jugadores de la NFL (a) Histograma (b) Histograma suavizado ■

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

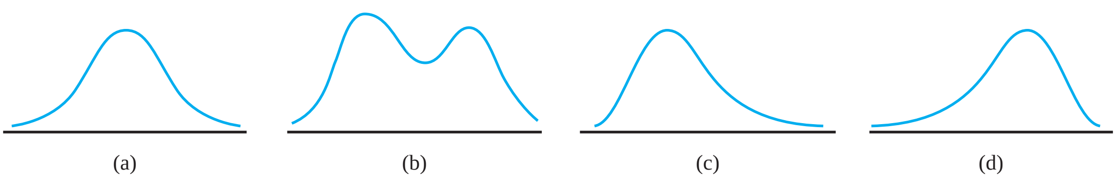


Figura 1.12 Histogramas suavizados: (a) unimodal simétrico; (b) bimodal; (c) positivamente asimétrico y (d) negativamente asimétrico

Datos cualitativos

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

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

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

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

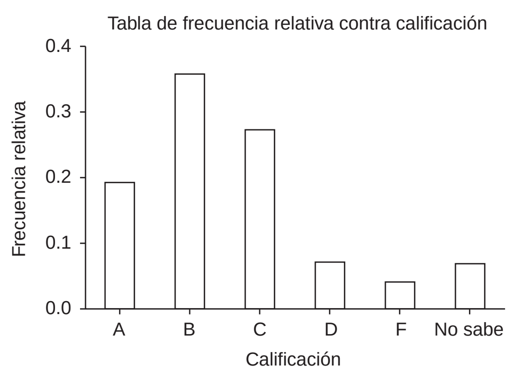


Figura 1.13 Histograma Minitab de los datos de la calificación

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

Datos multivariantes

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

EJERCICIOS Sección 1.2 (10–32)

10. Considere los datos de resistencia de las vigas del ejemplo 1.2.

a. Construya una gráfica de tallos y hojas de los datos. ¿Cuál parece ser el valor de resistencia representativo? ¿Parecen estar las observaciones altamente concentradas en torno al valor representativo o algo dispersas?

b. ¿Parece ser la gráfica razonablemente simétrica en torno a un valor representativo o describiría su forma de otra manera?

c. ¿Parece haber algunos valores de resistencia extremos?

d. ¿Qué proporción de las observaciones de resistencia en esta muestra exceden de 10 MPa?

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

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

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

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

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

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

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

a. Construya una gráfica de tallos y hojas de los datos eliminando (truncando) los dígitos de décimos y luego repitiendo cada valor de tallo cinco veces (una vez para las hojas 1 y 2, una segunda vez para las hojas 3 y 4, etc.). ¿Por qué es relativamente fácil identificar un valor de resistencia representativo?

b. Construya un histograma utilizando clases de ancho igual con la primera clase que tiene un límite inferior de 122 y un límite superior de 124. En seguida comente sobre cualquier característica interesante del histograma.

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

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

a. Construya una gráfica de tallos y hojas de los datos.

b. ¿Cuál es una velocidad de flujo o gasto típico o representativo?

c. ¿Parece estar la gráfica altamente concentrada o dispersa?

d. ¿Es la distribución de valores razonablemente simétrica? Si no, ¿cómo describiría el alejamiento de la simetría?

e. ¿Describiría alguna observación como alejada del resto de los datos (un valor extremo)?

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

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

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

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

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

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

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

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

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

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

Número de artículos	1	2	3	4	5	6	7	8	
Frecuencia	784	204	127	50	33	28	19	19	
Número de artículos	9	10	11	12	13	14	15	16	17
Frecuencia	6	7	6	7	4	4	5	3	3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

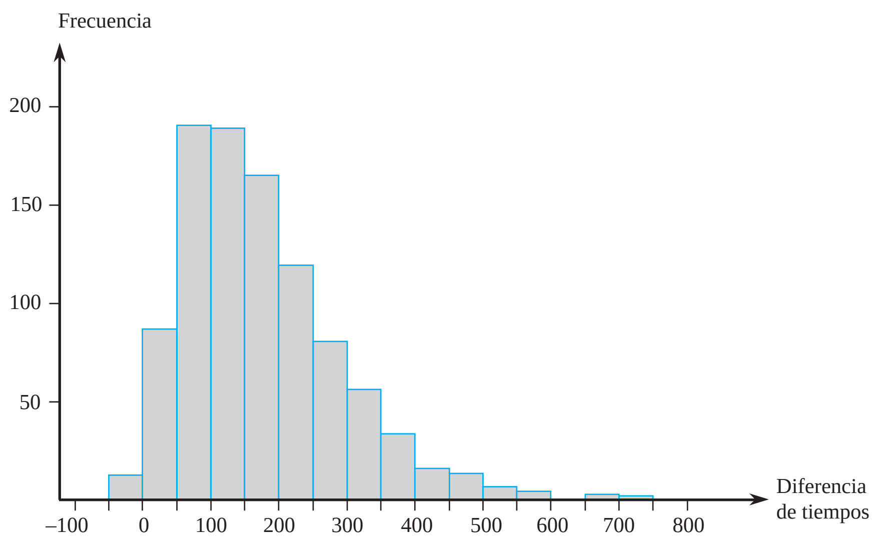
Construya un histograma y comente sobre las características interesantes.

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

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

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

Histograma para el ejercicio 22



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

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

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

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

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

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

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

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

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

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

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

Describa los datos y comente sobre cualquier característica importante.

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

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

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

31. La **frecuencia acumulativa** y la frecuencia relativa acumulativa de un intervalo de clase particular son la suma de frecuencias y frecuencias relativas, respectivamente, del intervalo y todos los intervalos que quedan debajo de él. Si, por ejemplo, existen cuatro intervalos con frecuencias 9, 16, 13 y 12, entonces las frecuencias acumulativas son 9, 25, 38 y 50 y las frecuencias relativas acumulativas son .18, .50, .76 y 1.00. Calcule las frecuencias acumulativas y las frecuencias relativas acumulativas de los datos del ejercicio 24.

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

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

a. Construya un histograma de frecuencia relativa y comente sobre las características interesantes.

b. ¿Qué proporción de cargas de fuego son menores que 600? ¿Por lo menos de 1200?

c. ¿Qué proporción de las cargas está entre 600 y 1200?

1.3 Medidas de ubicación

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

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

La media

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

DEFINICIÓN

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

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

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

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

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

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

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

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

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

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

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

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

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

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

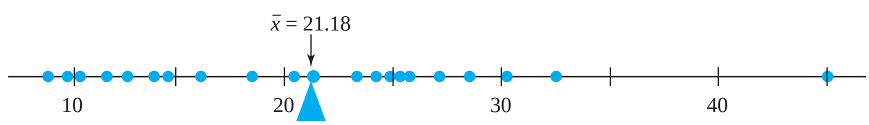


Figura 1.15 La media es el punto de equilibrio para un sistema de pesos

Así como $\bar{x}$ representa el valor promedio de las observaciones incluidas en una muestra, se puede calcular el promedio de todos los valores incluidos en la población. Este promedio se llama **media de la población** y está denotado por la letra griega μ. Cuando existen N valores en la población (una población finita), entonces μ = (suma de los N valores de población)/N. En los capítulos 3 y 4, se dará una definición más general de μ que se aplica tanto a poblaciones finitas y (conceptualmente) infinitas. Así como $\bar{x}$ es una medida interesante e importante de la ubicación de la muestra, μ es una interesante e importante característica (con frecuencia la más importante) de una población. En los capítulos sobre inferencia estadística, se presentarán métodos basados en la media muestral para sacar conclusiones con respecto a una media de población. Por ejemplo, se podría utilizar la media muestral $\bar{x} = 21.18$ calculada en el ejemplo 1.14 como una *estimación puntual* (un solo número que es la "mejor" conjetura) de $\mu =$ la longitud de agrietamiento promedio verdadera de todas las probetas tratadas como se describe.

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

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

La mediana

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

DEFINICIÓN

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

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

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

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

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

He aquí una gráfica de puntos de los datos:

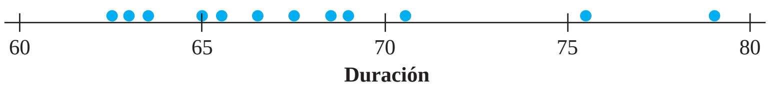


Figura 1.16 Gráfica de puntos de los datos para el ejemplo 1.14

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

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

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

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

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

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

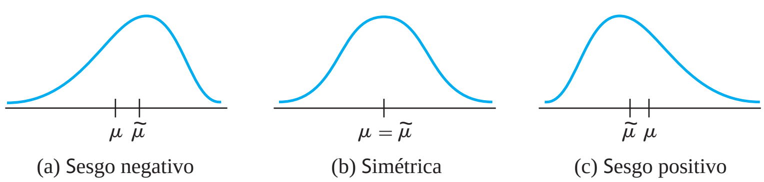


Figura 1.17 Tres formas diferentes de distribución de la población

Otras medidas de ubicación: cuartiles, percentiles y medias recortadas

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

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

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

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

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

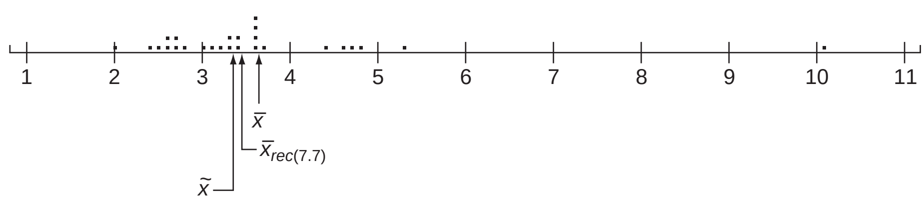


Figura 1.18 Gráfica de puntos del contenido de cobre para el ejemplo 1.16 ■

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

Datos categóricos y proporciones muestrales

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

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

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

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

EJERCICIOS Sección 1.3 (33–43)

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

590 815 575 608 350 1285 408 540 555 679

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

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

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

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

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

15.0 13.0 18.0 14.5 12.0 11.0 8.9 8.0

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

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

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

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

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

6.5 12.0 14.9 10.0 10.7 7.9 21.9 12.5 14.5 9.2

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

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

118.6 127.4 138.4 130.0 113.7 122.0 108.3
131.5 133.2

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

39. La propagación de grietas provocadas por fatiga en varias partes de un avión ha sido el tema de extensos estudios en años recientes. Los datos adjuntos se componen de vidas de propagación (horas de vuelo/10^4) para alcanzar un tamaño de agrietamiento dado en orificios para sujetadores utilizados en aviones militares ("Statistical Crack Propagation in Fastener Holes Under Spectrum Loading", *J. Aircraft*, 1983: 1028–1032):

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

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

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

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

S S F S S S F F S S

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

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

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

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

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

1.4 Medidas de variabilidad

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

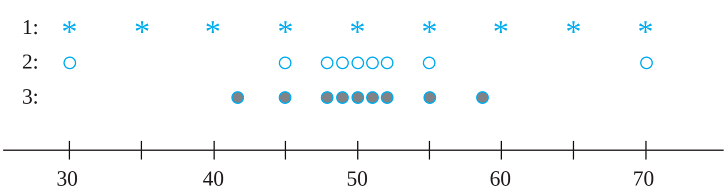


Figura 1.19 Muestras con medidas de centro idénticas pero diferentes cantidades de variabilidad

Medidas de variabilidad de datos muestrales

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

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

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

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

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

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

DEFINICIÓN

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

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

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

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

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

Ejemplo 1.17 El sitio web www.fueleconomy.gov contiene una gran cantidad de información acerca de las características del combustible de varios vehículos. Además de las calificaciones de millaje de la EPA, hay muchos vehículos para los que los usuarios han informado de sus propios valores de eficiencia de combustible (mpg). Considere la siguiente muestra de $n = 11$ eficiencias para el Ford Focus 2009 equipado con transmisión automática (para este modelo, la EPA informa de una calificación general de 27 mpg-24 mpg en ciudad y 33 mpg en carretera):

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

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

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

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

Motivación para s^2

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

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

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

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

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

Una fórmula para calcular s^2

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

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

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

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

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

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

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

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

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

Por tanto, el numerador de la varianza muestral es

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

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

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

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

PROPOSICIÓN

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

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

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

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

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

Gráficas de caja

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

DEFINICIÓN

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

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

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

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

x_i más pequeñas   cuarto inferior   mediana   cuarto superior x_i más grandes

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

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

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

El resumen de cinco números es como sigue:

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

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

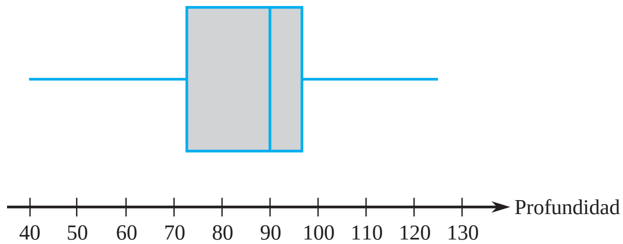


Figura 1.20 Gráfica de caja para los datos de corrosión

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

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

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

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

Gráficas de caja que muestran valores apartados

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

DEFINICIÓN

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

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

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

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

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

El resumen de las cantidades pertinentes es

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

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

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

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

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

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

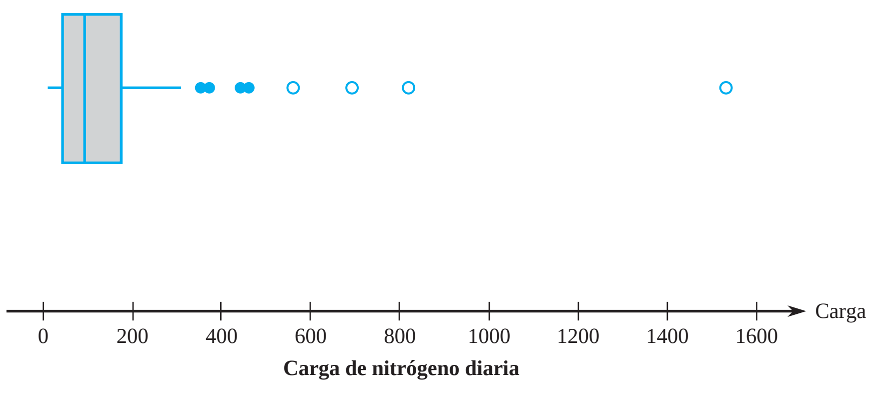


Figura 1.22 Gráfica de caja de los datos de la carga de nitrógeno mostrando los valores apartados moderados y extremos ■

Gráficas de caja comparativas

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

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

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

Figura 1.23 Gráfica de tallos y hojas para el ejemplo 1.21

El resumen de cantidades numéricas es el siguiente:

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

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

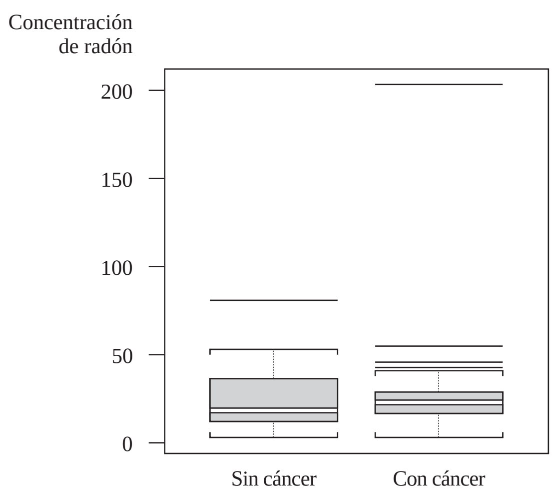


Figura 1.24 Gráfica de caja para el ejemplo 1.21, tomada de S-plus

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

EJERCICIOS Sección 1.4 (44–61)

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

29.5 49.3 30.6 28.2 28.0 26.3 33.9 29.4 23.5 31.6

Calcule lo siguiente

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

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

116.4 115.9 114.6 115.2 115.8

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

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

2781 2900 3013 2856 2888

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

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

87 93 96 98 105 114 128 131 142 168

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

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

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

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

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

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

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

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

50. En 1997 una mujer demandó a un fabricante de teclados de computadora y lo acusó de que sus repetidas lesiones por esfuerzo eran provocadas por el teclado (*Genessy v. Digital Equipment Corp.*). El jurado le adjudicó $3.5 millones por el dolor y sufrimiento pero la corte anuló dicha adjudicación por considerarla una compensación irrazonable. Al hacer esta determinación, la corte identificó un grupo "normativo" de 27 casos similares y especificó que una adjudicación razonable estaría dentro de dos desviaciones estándar de la media de las adjudicaciones en los 27 casos. Las 27 adjudicaciones fueron (en el rango de los $1000) 37, 60, 75, 115, 135, 140, 149, 150, 238, 290, 340, 410, 600, 750, 750, 750, 1050, 1100, 1139, 1150, 1200, 1200, 1250, 1576, 1700, 1825 y 2000, con las cuales $\sum x_i = 20{,}179$, $\sum x_i^2 = 24{,}657{,}511$. ¿Cuál es la cantidad máxima posible que podría ser adjudicada conforme a la regla de dos desviaciones estándar?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

mediana = 202.2 cuarto inferior = 196.0
cuarto superior = 216.8

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

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

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

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

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

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

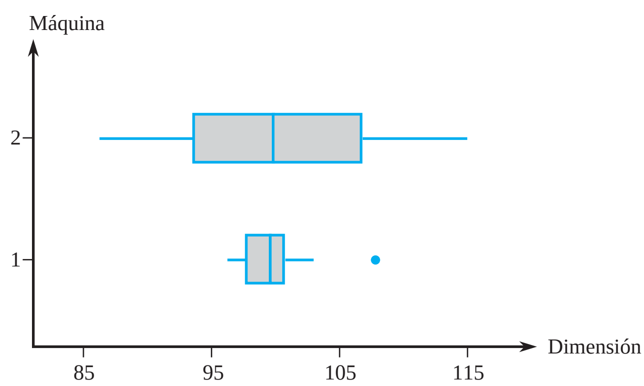
60. Se obtuvieron observaciones de resistencia al estallamiento (lb/pulg^2) con pruebas tanto con soldaduras de cierre de tobera como con soldaduras para tobera de envases de producción ("Proper Procedures Are the Key to Welding Radioactive Waste Cannisters", *Welding J.*, agosto de 1997: 61–67).

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

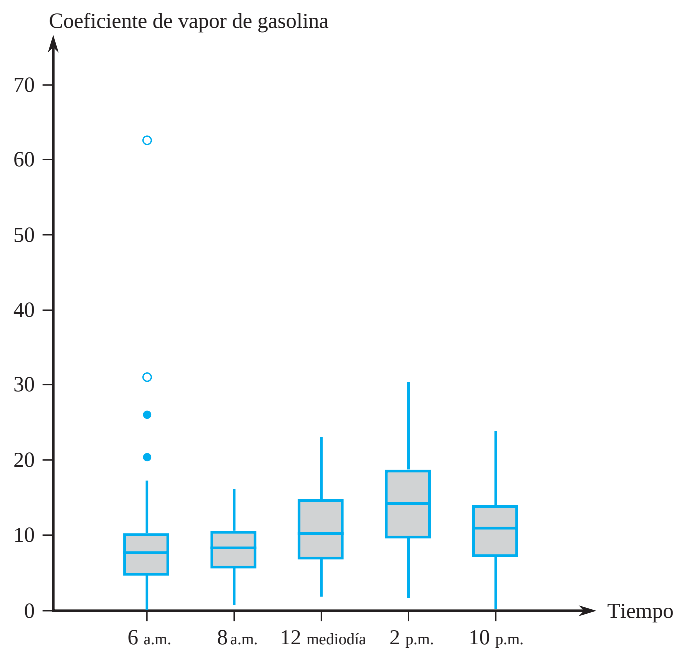
Construya una gráfica de caja comparativa y comente sobre las características interesantes (el artículo citado no incluía tal gráfica, pero los autores comentaron que habían visto una.)

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

Gráfica de caja comparativa para el ejercicio 58



Gráfica de caja comparativa para el ejercicio 61



EJERCICIOS SUPLEMENTARIOS (62–83)

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

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

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

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

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

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

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

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

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

b. Construya una gráfica de caja comparativa de los datos y comente las características interesantes.

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

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

a. Construya un histograma basado en frecuencias relativas y comente sobre cualesquiera características interesantes.

b. ¿Qué proporción de las observaciones de resistencia son por lo menos de 85? ¿Menores que 95?

c. Aproximadamente, ¿qué proporción de las observaciones son menores que 90?

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

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

a. ¿Parecen ser similares las concentraciones iniciales de Se en las muestras de suplemento y en las de control? Use varias técnicas de este capítulo para resumir los datos y responder la pregunta planteada.

b. De nuevo use métodos de este capítulo para resumir los datos y luego describa cómo los valores de concentración de Se finales en el grupo de tratamiento difieren de aquellos en el grupo de control.

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

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

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

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

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

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

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

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

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

Sujeto	1	2	3	4	5	6	7
Peso (x)	14.6	14.4	19.5	24.3	16.3	22.1	23.0
Caminadora (y)	11.3	5.3	9.1	15.2	10.1	19.6	20.8

Sujeto	8	9	10	11	12	13	14	15
Peso (x)	18.7	19.0	17.0	19.1	19.6	23.2	18.5	15.9
Caminadora (y)	10.3	10.3	2.6	16.6	22.4	23.6	12.6	4.4

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

a. Construya una gráfica de caja comparativa y comente sobre las similitudes y diferencias.

b. Construya una gráfica de puntos comparativa (una gráfica de puntos de cada muestra con una escala común). Comente sobre las similitudes y diferencias.

c. ¿Da la gráfica de caja comparativa del inciso (a) una evaluación informativa de similitudes y diferencias? Explique su razonamiento.

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

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

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

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

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

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

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

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

a. Sea $y_i = x_i - \bar{x}$ con $i = 1, \ldots, n$. ¿Cómo se comparan los valores de s^2 y s de las y_i con los valores correspondientes de las x_i? Explique.

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

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

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

b. Demuestre que

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

de modo que s_{n+1}^2 pueda ser calculada con x_{n+1}, $\bar{x}_n$ y s_n^2.

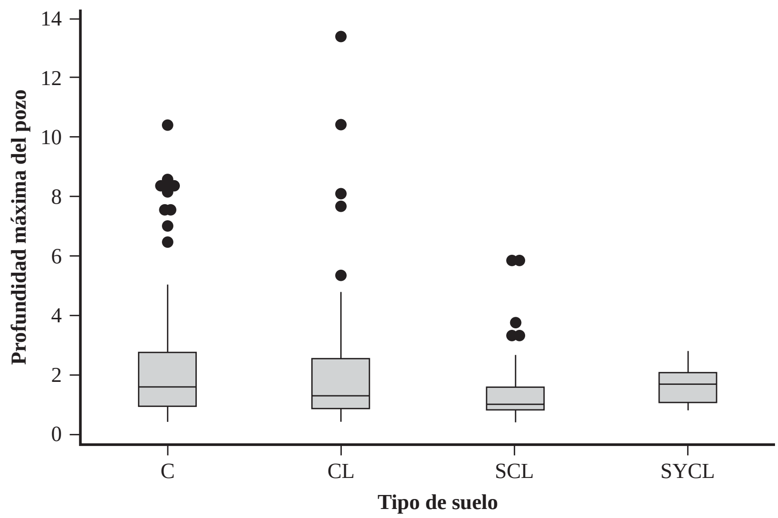
c. Suponga que una muestra de 15 torzales de hilo para telas dio por resultado una media muestral del alargamiento del hilo de 12.58 mm y una desviación estándar muestral de .512 mm. Una 16ava torzal resulta en un valor de alargamiento de 11.8. ¿Cuáles son los valores de la media muestral y la desviación estándar muestral de las 16 observaciones de alargamiento?

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

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

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

Gráfica de caja comparativa para el ejercicio 77



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

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

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

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

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

82. Los datos muestrales $x_1, x_2, \ldots, x_n$ en ocasiones representan una **serie de tiempo**, donde $x_t =$ el valor observado de una variable de respuesta x en el tiempo t. A menudo la serie observada muestra una gran cantidad de variación aleatoria, lo que dificulta estudiar el comportamiento a plazo más largo. En tales situaciones, es deseable producir una versión suavizada de la serie. Una técnica para hacerlo implica el **suavizamiento exponencial**. Se elige el valor de una constante de suavizamiento α $(0 < \alpha < 1)$. Luego con $\bar{x}_t =$ valor suavizado en el tiempo t, se hace $\bar{x}_1 = x_1$ y para $t = 2, 3, \ldots, n$, $\bar{x}_t = \alpha x_t + (1 - \alpha)\bar{x}_{t-1}$.
a. Considere la siguiente serie de tiempo en la cual $x_t =$ temperatura (°F) del efluente en una planta de tratamiento de aguas negras en el día t: 47, 54, 53, 50, 46, 46, 47, 50, 51, 50, 46, 52, 50, 50. Trace cada x_t contra t en un sistema de coordenadas de dos dimensiones (una gráfica de tiempo-serie). ¿Parece haber algún patrón?
b. Calcule las $\bar{x}_t$ con $\alpha = .1$. Repita con $\alpha = .5$. ¿Qué valor de α da una serie $\bar{x}_t$ más atenuada?
c. Sustituya $\bar{x}_{t-1} = \alpha x_{t-1} + (1 - \alpha)\bar{x}_{t-2}$ en el miembro de la derecha de la expresión para $\bar{x}_t$, acto seguido sustituya $\bar{x}_{t-2}$ en función de x_{t-2} y $\bar{x}_{t-3}$, y así sucesivamente. ¿De cuántos de los valores $x_t, x_{t-1}, \ldots, x_1$ depende $\bar{x}_t$. ¿Qué le sucede al coeficiente de x_{t-k} conforme k se incrementa?
d. Remítase al inciso (c). Si t es grande, ¿qué tan sensible es $\bar{x}_t$ a la inicialización $\bar{x}_1 = x_1$? Explique.

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

83. Considere las observaciones numéricas $x_1, \ldots, x_n$. Con frecuencia interesa saber si las x_i están (por lo menos en forma aproximada) simétricamente distribuidas en torno a algún mismo valor. Si n es por lo menos grande de manera moderada, el grado de simetría puede ser valorado con una gráfica de tallo y hojas o un histograma. Sin embargo, si n no es muy grande, las gráficas mencionadas no son informativas en particular. Considere la siguiente alternativa. Que y_1 denote la x_i más pequeña, y_2 la segunda x_i más pequeña, y así sucesivamente. Luego grafique los siguientes pares como puntos en un sistema de coordenadas de dos dimensiones $(y_n - \tilde{x}, \tilde{x} - y_1)$, $(y_{n-1} - \tilde{x}, \tilde{x} - y_2)$, $(y_{n-2} - \tilde{x}, \tilde{x} - y_3), \ldots$ Existen $n/2$ puntos cuando n es par y $(n - 1)/2$ cuando n es impar.
a. ¿Qué apariencia tiene esta gráfica cuando la simetría en los datos es perfecta? ¿Qué apariencia tiene cuando las observaciones se alargan más sobre la mediana que debajo de ella (una larga cola superior)?
b. Los datos adjuntos sobre cantidad de lluvia (acres-pies) producida por 26 nubes bombardeadas se tomaron del artículo "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification" (*Technometrics*, 1975: 161–166). Construya la gráfica y comente sobre el grado de simetría o la naturaleza del alejamiento de la misma.

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

Bibliografía

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

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

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

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

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

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

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

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

2 Probabilidad

INTRODUCCIÓN

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

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

2.1 Espacios muestrales y eventos

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

El espacio muestral de un experimento

DEFINICIÓN

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

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

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

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

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

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

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

■

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

Eventos

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

DEFINICIÓN

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

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

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

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

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

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

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

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

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

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

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

Ejemplo 2.7 (continuación del ejemplo 2.4) El espacio muestral del experimento del programa de compilación contiene un número infinito de resultados, por lo que existe un número infinito de eventos simples. Los eventos compuestos incluyen

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

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

Algunas relaciones de la teoría de conjuntos

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

DEFINICIÓN

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

Ejemplo 2.8 (continuación del ejemplo 2.3) En el experimento en el cual se observa el número de bombas en uso en una sola gasolinera de seis bombas, sea $A = \{0, 1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ y $C = \{1, 3, 5\}$. Entonces

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

Ejemplo 2.9 (continuación del ejemplo 2.4) En el experimento de compilación de programas defina A, B y C como

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

Entonces

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

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

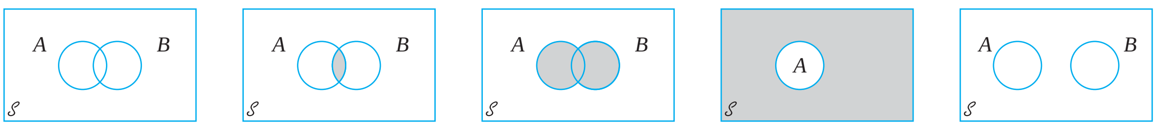
DEFINICIÓN

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

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

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

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



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

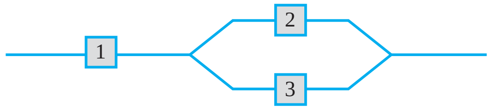
Figura 2.1 Diagramas de Venn

EJERCICIOS Sección 2.1 (1–10)

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

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

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



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

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

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

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

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

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

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

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

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

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

2.2 Axiomas, interpretaciones, y propiedades de la probabilidad

Dados un experimento y un espacio muestral $\mathcal{S}$, *el objetivo de la probabilidad es asignar a cada evento A un número P(A), llamado la probabilidad del evento A, el cual dará una medida precisa de la oportunidad de que A ocurrirá*. Para garantizar que las asignaciones serán consistentes con las nociones intuitivas de la probabilidad, todas las asignaciones deberán satisfacer los siguientes axiomas (propiedades básicas) de probabilidad.

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

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

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

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

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

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

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

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

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

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

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

como se deseaba. ■

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

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

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

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

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

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

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

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

Interpretación de probabilidad

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

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

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

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

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

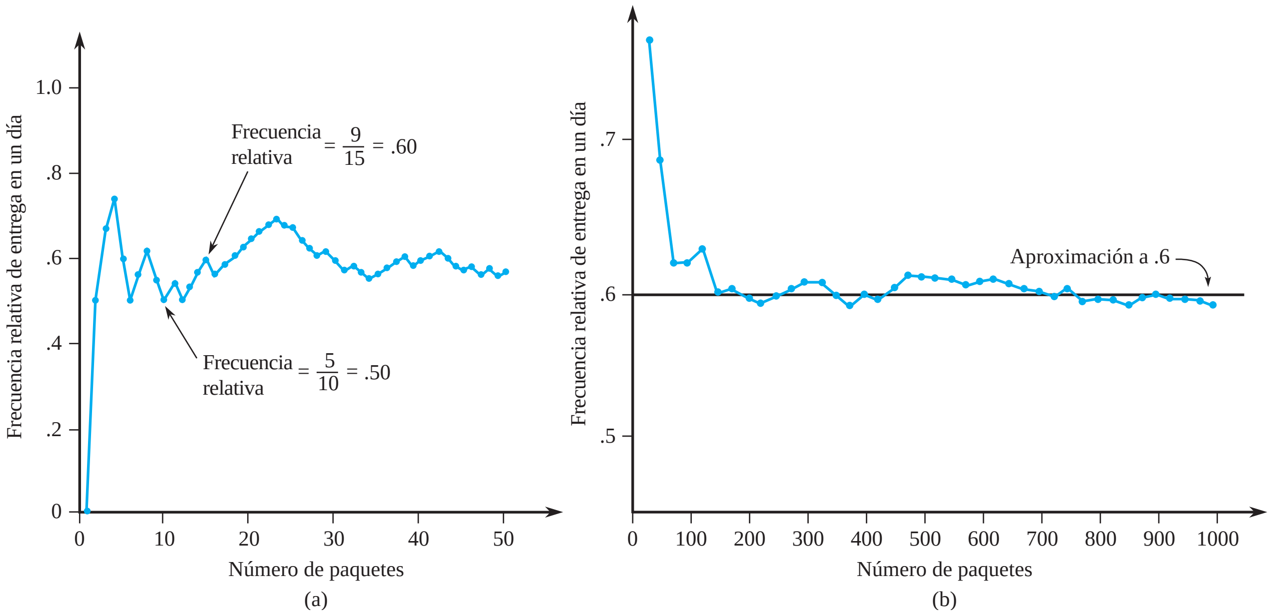


Figura 2.2 Comportamiento de la frecuencia relativa (a) fluctuación inicial (b) estabilización a largo plazo.

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

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

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

Más propiedades de probabilidad

PROPOSICIÓN

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

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

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

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

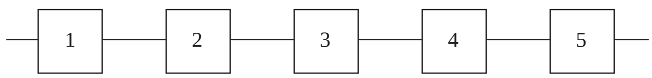


Figura 2.3 Sistema de cinco componentes conectados en serie

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

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

PROPOSICIÓN

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

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

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

PROPOSICIÓN

Para dos eventos cualesquiera A y B,

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

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

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

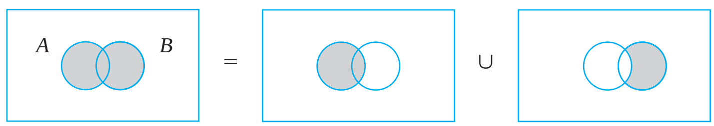


Figura 2.4 Representación de A ∪ B como la unión de dos eventos excluyentes ■

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

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

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

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

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

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

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

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

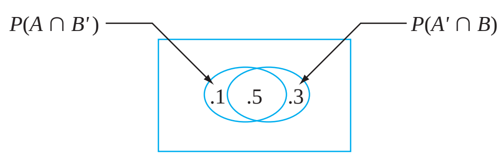


Figura 2.5 Probabilidades para el ejemplo 2.14 ■

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

Para tres eventos cualesquiera A, B y C,

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

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

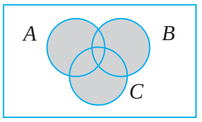


Figura 2.6 $A \cup B \cup C$

Determinación sistemática de probabilidades

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

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

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

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

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

Resultados igualmente probables

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

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

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

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

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

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

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

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

■

EJERCICIOS Sección 2.2 (11–28)

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

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

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

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

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

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

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

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

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

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

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

a. $A_1 \cup A_2$

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

c. $A_1 \cup A_2 \cup A_3$

d. $A_1' \cap A_2' \cap A_3'$

e. $A_1' \cap A_2' \cap A_3$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.3 Técnicas de conteo

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

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

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

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

La regla de producto para pares ordenados

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

PROPOSICIÓN

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

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

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

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

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

En muchos problemas de conteo y probabilidad se puede utilizar una configuración conocida como **diagrama de árbol** para representar pictóricamente todas las posibilidades. El diagrama de árbol asociado con el ejemplo 2.18 aparece en la figura 2.7. Partiendo de un punto localizado en el lado izquierdo del diagrama, por cada posible primer elemento de un par emana un segmento de línea recta hacia la derecha. Cada una de estas líneas se conoce como rama de primera generación. Ahora para cualquier rama de primera generación se construye otro segmento de línea que emana de la punta de la rama por cada posible opción de un segundo elemento del par. Cada segmento de línea es una rama de segunda generación. Como existen cuatro obstetras, existen cuatro ramas de primera generación y tres pediatras por cada obstetra, resultan tres ramas de segunda generación que emanan de cada rama de primera generación.

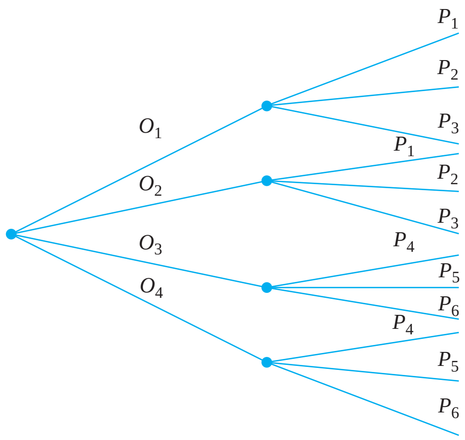


Figura 2.7 Diagrama de árbol para el ejemplo 2.18

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

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

Una regla de producto más general

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

Regla de producto para *k*-tuplas

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

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

Ejemplo 2.19 (continuación del ejemplo 2.17) Suponga que el trabajo de remodelación de la casa implica adquirir primero varios utensilios de cocina. Se adquirirán en la misma tienda y hay cinco tiendas en el área. Con las tiendas denotadas por $D_1, \ldots, D_5$, existen $N = n_1 n_2 n_3 = (5)(12)(9) = 540$ 3-tuplas de la forma (D_i, P_j, Q_k), así que existen 540 formas de elegir primero una tienda, luego un contratista de fontanería y finalmente un contratista electricista. ■

Ejemplo 2.20 (continuación del ejemplo 2.18) Si cada clínica tiene dos especialistas en medicina interna y dos médicos generales, existen $n_1 n_2 n_3 n_4 = (4)(3)(3)(2) = 72$ formas de seleccionar un doctor de cada tipo, de tal suerte que todos los doctores practiquen en la misma clínica. ■

Permutaciones y combinaciones

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

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

DEFINICIÓN

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

El número de permutaciones se determina utilizando la primera regla de conteo para k-tuplas. Supóngase, por ejemplo, que un colegio de ingeniería tiene siete departamentos, denotados por a, b, c, d, e, f y g. Cada departamento tiene un representante en el consejo de estudiantes del colegio. De estos siete representantes, uno tiene que ser elegido como presidente, otro como vicepresidente y un tercero como secretario. ¿Cuántas maneras de seleccionar los tres oficiales existen? Es decir, ¿cuántas permutaciones de tamaño 3 pueden ser formadas con los 7 representantes? Para responder esta pregunta, habrá que pensar en formar una terna (3-tupla) en el cual el primer elemento es el presidente, el segundo es el vicepresidente y el tercero es el secretario. Una terna es (a, g, b), otro es (b, g, a) y otro más es (d, f, b). Ahora bien, el presidente puede ser seleccionado en cualesquiera de $n_1 = 7$ formas. Por cada forma de seleccionar el presidente, existen $n_2 = 6$ formas de seleccionar el vicepresidente y por consiguiente $7 \times 6 = 42$ (pares de presidente, vicepresidente). Por último, por cada forma de seleccionar un presidente y vicepresidente, existen $n_3 = 5$ formas de seleccionar el secretario. Esto da

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

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

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

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

Más generalmente,

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

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

PROPOSICIÓN

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

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

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

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

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

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

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

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

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

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

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

PROPOSICIÓN

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

■

EJERCICIOS Sección 2.3 (29–44)

29. Con fecha de abril de 2006, aproximadamente 50 millones de nombres de dominio web.com fueron registrados (p. ej., yahoo.com).

a. ¿Cuántos nombres de dominio compuestos de exactamente dos letras en sucesión pueden ser formados? ¿Cuántos nombres de dominio de dos letras existen si como caracteres se permiten dígitos y letras? [*Nota*: una longitud de carácter de tres o más ahora es obligatoria.]

b. ¿Cuántos nombres de dominio existen compuestos de tres letras en secuencia? ¿Cuántos de esta longitud existen si se permiten letras o dígitos? [*Nota*: en la actualidad todos están utilizados.]

c. Responda las preguntas hechas en (b) para secuencias de cuatro caracteres.

d. Con fecha de abril de 2006, 97,786 de las secuencias de cuatro caracteres utilizando o letras o dígitos aún no habían sido reclamadas. Si se elige un nombre de cuatro caracteres al azar, ¿cuál es la probabilidad de que ya tenga dueño?

30. Un amigo mío va a ofrecer una fiesta. Sus existencias actuales de vino incluye 8 botellas de zinfandel, 10 de merlot y 12 de cabernet (él sólo bebe vino tinto), todos de diferentes fábricas vinícolas.

a. Si desea servir 3 botellas de zinfandel y el orden de servicio es importante, ¿cuántas formas existen de hacerlo?

b. Si 6 botellas de vino tienen que ser seleccionadas al azar de las 30 para servirse, ¿cuántas formas existen de hacerlo?

c. Si se seleccionan al azar 6 botellas, ¿cuántas formas existen de obtener dos botellas de cada variedad?

d. Si se seleccionan 6 botellas al azar, ¿cuál es la probabilidad de que el resultado sea dos botellas de cada variedad?

e. Si se eligen 6 botellas al azar, ¿cuál es la probabilidad de que todas ellas sean de la misma variedad?

31. a. Beethoven escribió 9 sinfonías y Mozart 27 conciertos para piano. Si el locutor de una estación de radio de una universidad desea transmitir primero una sinfonía de Beethoven y luego un concierto de Mozart, ¿de cuántas maneras puede hacerlo?

b. El gerente de la estación decide que en cada noche sucesiva (7 días a la semana), se tocará una sinfonía de Beethoven, seguida por un concierto para piano de Mozart, seguido por un cuarteto de cuerdas de Schubert (de los cuales existen 15). ¿Durante aproximadamente cuántos años se podría continuar con esta política antes de que exactamente el mismo programa se repitiera?

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

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

Un tablero de distribución en la tienda permite al cliente conectar cualquier selección de componentes (compuesta de uno de cada tipo). Use las reglas de producto para responder las siguientes preguntas.

a. ¿De cuántas maneras puede ser seleccionado un componente de cada tipo?

b. ¿De cuántas maneras pueden ser seleccionados los componentes si tanto el receptor como el reproductor de discos compactos tiene que ser Sony?

c. ¿De cuántas maneras pueden ser seleccionados los componentes si ninguno tiene que ser Sony?

d. ¿De cuántas maneras se puede hacer una selección si por lo menos se tiene que incluir un componente Sony?

e. Si alguien mueve los interruptores en el tablero de distribución completamente al azar, ¿cuál es la probabilidad de que el sistema seleccionado contenga por lo menos un componente Sony? ¿Exactamente un componente Sony?

33. De nuevo considere el equipo de ligas pequeñas que tiene 15 jugadores en su plantel.

a. ¿Cuántas formas existen de seleccionar 9 jugadores para la alineación inicial?

b. ¿Cuántas formas existen de seleccionar 9 jugadores para la alineación inicial y un orden al bat de los nueve inicialistas?

c. Suponga que 5 de los 15 jugadores son zurdos. ¿Cuántas formas existen de seleccionar 3 jardineros zurdos y tener las otras 6 posiciones ocupadas por jugadores derechos?

34. Fallas del teclado de computadora pueden ser atribuidas a defectos eléctricos o mecánicos. Un taller de reparación actualmente cuenta con 25 teclados averiados, de los cuales 6 tienen defectos eléctricos y 19 tienen defectos mecánicos.

a. ¿Cuántas maneras hay de seleccionar al azar cinco de estos teclados para una inspección completa (sin tener en cuenta el orden)?

b. ¿De cuántas maneras puede seleccionarse una muestra de 5 teclados de manera que sólo dos tengan un defecto eléctrico?

c. Si una muestra de 5 teclados se selecciona al azar, ¿cuál es la probabilidad de que al menos 4 de éstos tengan un defecto mecánico?

35. Una empresa de producción emplea 20 trabajadores en el turno de día, 15 en el turno de tarde y 10 en el turno de medianoche. Un consultor de control de calidad va a seleccionar 6 de estos trabajadores para entrevistas a fondo. Suponga que la selección se hace de tal modo que cualquier grupo particular de 6 trabajadores tiene la misma oportunidad de ser seleccionado al igual que cualquier otro grupo (sacando 6 papelitos de entre 45 sin reemplazarlos).

a. ¿Cuántas selecciones resultarán en que los 6 trabajadores seleccionados provengan del turno de día? ¿Cuál es la probabilidad de que los 6 trabajadores seleccionados sean del turno de día?

b. ¿Cuál es la probabilidad de que los 6 trabajadores seleccionados sean del mismo turno?

c. ¿Cuál es la probabilidad de que por lo menos dos turnos diferentes estén representados entre los trabajadores seleccionados?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.4 Probabilidad condicional

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

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

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

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

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

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

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

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

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

■

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

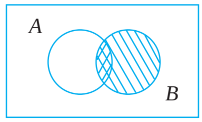


Figura 2.8 Motivación de la definición de probabilidad condicional

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

Definición de probabilidad condicional

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

DEFINICIÓN

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

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

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

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

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

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

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

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

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

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

La figura 2.9 ilustra las probabilidades pertinentes.

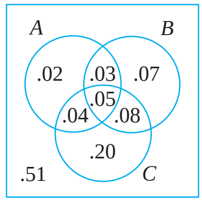


Figura 2.9 Diagrama de Venn para el ejemplo 2.26

Por lo tanto se tiene

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

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

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

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

y

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

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

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

La regla de multiplicación

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

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

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

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

$$\begin{aligned} P(\text{por lo menos tres individuos fueron examinados} \\ \text{para determinar su tipo de sangre}) &= P(A \cap B) \\ &= P(A|B) \cdot P(B) \\ &= \frac{2}{3} \cdot \frac{3}{4} = \frac{6}{12} \\ &= .5 \end{aligned}$$

La regla de multiplicación es más útil cuando los experimentos se componen de varias etapas en sucesión. El evento condicionante B describe entonces el resultado de la primera etapa y A el resultado de la segunda, de modo que $P(A|B)$, condicionada en lo que ocurra primero, a menudo será conocida. La regla es fácil de ser ampliada a experimentos que implican más de dos etapas. Por ejemplo,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3|A_1 \cap A_2) \cdot P(A_1 \cap A_2) \\ &= P(A_3|A_1 \cap A_2) \cdot P(A_2|A_1) \cdot P(A_1) \end{aligned} \tag{2.4}$$

donde A_1 ocurre primero, seguido por A_2 y finalmente A_3.

Ejemplo 2.28 Para el experimento de determinación de tipo de sangre del ejemplo 2.27,

$$\begin{aligned} P(\text{el tercer tipo es O+}) &= P(\text{el tercero es}|\text{el primero no es} \cap \text{el segundo no es}) \\ &\quad \cdot P(\text{el segundo no es}|\text{el primero no es}) \cdot P(\text{el primero no es}) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4} = .25 \end{aligned}$$

Cuando el experimento de interés se compone de una secuencia de varias etapas, es conveniente representarlas con un diagrama de árbol. Una vez que se tiene un diagrama de árbol apropiado, las probabilidades y las probabilidades condicionales pueden ser ingresadas en las diversas ramas; esto implicará el uso repetido de la regla de multiplicación.

Ejemplo 2.29 Una cadena de tiendas de video vende tres marcas diferentes de reproductores de DVD. De sus ventas de reproductores de DVD, 50% son de la marca 1 (la menos cara), 30% son de la marca 2 y 20% son de la marca 3. Cada fabricante ofrece 1 año de garantía en las partes y mano de obra. Se sabe que 25% de los reproductores de DVD de la marca 1 requieren trabajo de reparación dentro del periodo de garantía, mientras que los porcentajes correspondientes de las marcas 2 y 3 son 20% y 10%, respectivamente.

1. ¿Cuál es la probabilidad de que un comprador seleccionado al azar haya adquirido un reproductor de DVD marca 1 que necesitará reparación mientras se encuentra dentro de la garantía?
2. ¿Cuál es la probabilidad de que un comprador seleccionado al azar haya comprado un reproductor de DVD que necesitará reparación mientras se encuentra dentro de la garantía?
3. Si un cliente regresa a la tienda con un reproductor de DVD que necesita reparación dentro de la garantía, ¿cuál es la probabilidad de que sea un reproductor de DVD marca 1? ¿Un reproductor de DVD marca 2? ¿Un reproductor de DVD marca 3?

La primera etapa del problema implica un cliente que selecciona una de las tres marcas de reproductor de DVD. Sea A_i = {marca i adquirida}, con $i = 1, 2$, y 3. Entonces $P(A_1) = .50$, $P(A_2) = .30$ y $P(A_3) = .20$. Una vez que se selecciona una marca de reproductor de DVD, la segunda etapa implica observar si el reproductor de DVD seleccionado necesita reparación dentro de la garantía. Con B = {necesita reparación} y B' = {no necesita reparación}, la información dada implica que $P(B|A_1) = .25$, $P(B|A_2) = .20$ y $P(B|A_3) = .10$.

El diagrama de árbol que representa esta situación experimental se muestra en la figura 2.10. Las ramas iniciales corresponden a marcas diferentes de reproductores de DVD; hay dos ramas de segunda generación que emanan de la punta de cada rama inicial, una para "necesita reparación" y la otra para "no necesita reparación". La probabilidad $P(A_i)$ aparece en la rama i-ésima inicial, en tanto que las probabilidades condicionales $P(B|A_i)$ y $P(B'|A_i)$ aparecen en las ramas de la segunda generación. A la derecha de cada rama de segunda generación correspondiente a la ocurrencia de B, se muestra el producto de probabilidades en las ramas que conducen hacia fuera de dicho punto. Ésta es simplemente la regla de multiplicación en acción. La respuesta a la pregunta planteada en 1 es por lo tanto $P(A_1 \cap B) = P(B|A_1) \cdot P(A_1) = .125$. La respuesta a la pregunta 2 es

$$\begin{aligned} P(B) &= P[(\text{marca 1 y reparación}) \text{ o } (\text{marca 2 y reparación}) \text{ o } (\text{marca 3 y reparación})] \\ &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\ &= .125 + .060 + .020 = .205 \end{aligned}$$

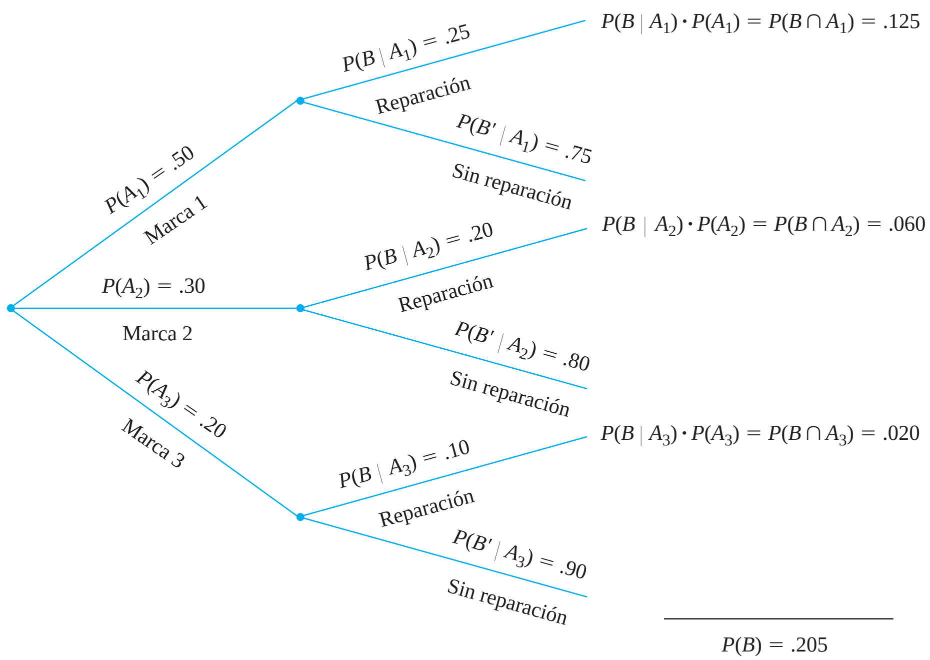


Figura 2.10 Diagrama de árbol para el ejemplo 2.29

Finalmente,

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.125}{.205} = .61$$

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{.060}{.205} = .29$$

y

$$P(A_3|B) = 1 - P(A_1|B) - P(A_2|B) = .10$$

La *probabilidad previa* o inicial de la marca 1 es .50. Una vez que se sabe que el reproductor de DVD seleccionado necesitaba reparación, la *probabilidad posterior* de la marca 1 se incrementa a .61. Esto se debe a que es más probable que los reproductores de DVD marca 1 necesiten reparación de garantía que las demás marcas. La probabilidad posterior de la marca 3 es $P(A_3 \mid B) = .10$, la cual es mucho menor que la probabilidad previa $P(A_3) = .20$. ■

Teorema de Bayes

El cálculo de una probabilidad posterior $P(A_j \mid B)$ a partir de probabilidades previas dadas $P(A_i)$ y probabilidades condicionales $P(B \mid A_i)$ ocupa una posición central en la probabilidad elemental. La regla general de dichos cálculos, los que en realidad son una aplicación simple de la regla de multiplicación, se remonta al reverendo Thomas Bayes, quien vivió en el siglo XVIII. Para formularla primero se requiere otro resultado. Recuérdese que los eventos $A_1, \ldots, A_k$ son mutuamente exclusivos si ninguno de los dos tiene resultados comunes. Los eventos son *exhaustivos* si un A_i debe ocurrir, de modo que $A_1 \cup \ldots \cup A_k = \mathcal{S}$.

Ley de probabilidad total

Sean $A_1, \ldots, A_k$ eventos mutuamente exclusivos y exhaustivos. Entonces para cualquier otro evento B,

$$\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k) \\ &= \sum_{i=1}^{k} P(B \mid A_i)P(A_i) \qquad (2.5) \end{aligned}$$

Comprobación Como los eventos A_i son mutuamente exclusivos y exhaustivos, si B ocurre debe ser en forma conjunta con uno de los eventos A_i de manera exacta. Es decir, $B = (A_1 \cap B) \cup \ldots \cup (A_k \cap B)$, donde los eventos $(A_i \cap B)$ son mutuamente exclusivos. Esta "partición de B" se ilustra en la figura 2.11. Por lo tanto

$$P(B) = \sum_{i=1}^{k} P(A_i \cap B) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$

como se deseaba.

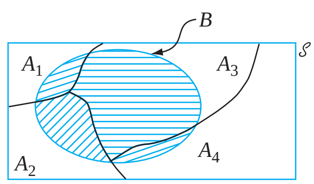


Figura 2.11 Partición de B entre las A_i' mutuamente excluyentes y exhaustivas. ■

Ejemplo 2.30 Una persona tiene 3 cuentas de correo electrónico diferentes. La mayoría de sus mensajes, el 70%, entra en la cuenta # 1, mientras que el 20% entra en la cuenta # 2 y el 10% restante en la cuenta # 3. De los mensajes en la cuenta # 1, sólo el 1% son spam, mientras que los porcentajes correspondientes a las cuentas # 2 y # 3 son 2% y 5%, respectivamente. ¿Cuál es la probabilidad de que un mensaje spam sea seleccionado al azar?

Para responder a esta pregunta, primero vamos a establecer una notación:

$$A_i = \{\text{el mensaje es de la cuenta } \# \, i\} \text{ para } i = 1, 2, 3, \quad B = \{\text{el mensaje es spam}\}$$

Entonces, los porcentajes dados implican que

$$P(A_1) = .70,\ P(A_2) = .20,\ P(A_3) = .10$$
$$P(B|A_1) = .01,\ P(B|A_2) = .02,\ P(B|A_3) = .05$$

Ahora bien, es simplemente una cuestión de sustituir en la ecuación de la ley de probabilidad total:

$$P(B) = (.01)(.70) + (.02)(.20) + (.05)(.10) = .016$$

A largo plazo, el 1.6% de los mensajes de esta persona serán spam. ■

Teorema de Bayes

Sean $A_1, A_2, \ldots, A_k$ un conjunto de eventos mutuamente exclusivos y exhaustivos con probabilidades *previas* $P(A_i)$ $(i = 1, \ldots, k)$. Entonces para cualquier otro evento B para el cual $P(B) > 0$, la probabilidad *posterior* de A_j dado que B ha ocurrido es

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)} \qquad j = 1, \ldots, k \tag{2.6}$$

La transición de la segunda a la tercera expresión en (2.6) se apoya en el uso de la regla de multiplicación en el numerador y la ley de probabilidad total en el denominador. La proliferación de eventos y subíndices en (2.6) puede ser un poco intimidante para los recién llegados a la probabilidad. Mientras existan relativamente pocos eventos en la repartición, se puede utilizar un diagrama de árbol (como en el ejemplo 2.29) como base para calcular probabilidades posteriores sin jamás referirse de manera explícita al teorema de Bayes.

Ejemplo 2.31 *Incidencia de una enfermedad rara*. Sólo 1 de 1000 adultos padece una enfermedad rara para la cual se ha creado una prueba de diagnóstico. La prueba es tal que cuando un individuo en realidad tiene la enfermedad, un resultado positivo se presentará en 99% de las veces mientras que en individuos sin enfermedad el examen será positivo sólo el 2% de las veces. Si se somete a prueba un individuo seleccionado al azar y el resultado es positivo, ¿cuál es la probabilidad de que el individuo tenga la enfermedad?

Para utilizar el teorema de Bayes, sea A_1 = el individuo tiene la enfermedad, A_2 = el individuo no tiene la enfermedad y B = resultado de prueba positivo. Entonces $P(A_1) = .001$, $P(A_2) = .999$, $P(B|A_1) = .99$ y $P(B|A_2) = .02$. El diagrama de árbol para este problema aparece en la figura 2.12.

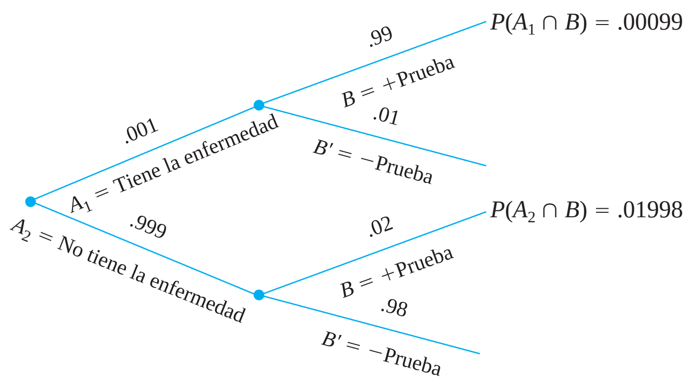


Figure 2.12 Diagrama de árbol para el problema de una enfermedad rara

Junto a cada rama correspondiente a un resultado positivo de prueba, la regla de multiplicación da las probabilidades anotadas. Por consiguiente, $P(B) = .00099 + .01998 = .02097$, a partir de la cual se tiene

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.00099}{.02097} = .047$$

Este resultado parece contraintuitivo; la prueba de diagnóstico parece tan precisa que es altamente probable que alguien con un resultado positivo de prueba tenga la enfermedad, mientras que la probabilidad condicional calculada es de sólo .047. Sin embargo, como la enfermedad es rara y la prueba es sólo moderadamente confiable, surgen más resultados positivos de prueba a causa de errores y no de individuos enfermos. La probabilidad de tener la enfermedad se ha incrementado por un factor de multiplicación de 47 (desde la probabilidad previa de .001 hasta la probabilidad posterior de .047); pero para incrementar aún más la probabilidad posterior, se requiere una prueba de diagnóstico con tasas de error mucho más pequeñas. ■

EJERCICIOS Sección 2.4 (45–69)

45. La población de un país particular se compone de tres grupos étnicos. Cada individuo pertenece a uno de los cuatro grupos sanguíneos principales. La *tabla de probabilidad conjunta* anexa da la proporción de individuos en las diversas combinaciones de grupo étnico–grupo sanguíneo.

		Grupo sanguíneo			
		O	**A**	**B**	**AB**
Grupo étnico	**1**	.082	.106	.008	.004
	2	.135	.141	.018	.006
	3	.215	.200	.065	.020

Suponga que se selecciona un individuo al azar de la población y que los eventos se definen como A = {tipo A seleccionado}, B = {tipo B seleccionado} y C = {grupo étnico 3 seleccionado}.

a. Calcule $P(A)$, $P(C)$ y $P(A \cap C)$.

b. Calcule tanto $P(A \mid C)$ y $P(C \mid A)$ y explique en contexto lo que cada una de estas probabilidades representa.

c. Si el individuo seleccionado no tiene sangre de tipo B, ¿cuál es la probabilidad de que él o ella pertenezcan al grupo étnico 1?

46. Suponga que un individuo es seleccionado al azar de la población de todos los adultos varones que viven en Estados Unidos. Sea A el evento en que el individuo seleccionado tiene una estatura de más de 6 pies, y sea B el evento en que el individuo seleccionado es un jugador profesional de basquetbol. ¿Cuál piensa que es más grande, $P(A \mid B)$ o $P(B \mid A)$? ¿Por qué?

47. Regrese al escenario de la tarjeta de crédito del ejercicio 12 (sección 2.2), donde A = {Visa}, B = {MasterCard}, $P(A) = .5$, $P(B) = .4$ y $P(A \cap B) = .25$. Calcule e interprete cada una de las siguientes probabilidades (un diagrama de Venn podría ayudar).

a. $P(B \mid A)$ **b.** $P(B' \mid A)$

c. $P(A \mid B)$ **d.** $P(A' \mid B)$

e. Dado que el individuo seleccionado tiene por lo menos una tarjeta, ¿cuál es la probabilidad de que él o ella tengan una tarjeta Visa?

48. Reconsidere la situación del sistema defectuoso descrito en el ejercicio 26 (sección 2.2).

a. Dado que el sistema tiene un defecto de tipo 1, ¿cuál es la probabilidad de que tenga un defecto de tipo 2?

b. Dado que el sistema tiene un defecto de tipo 1, ¿cuál es la probabilidad de que tenga los tres tipos de defecto?

c. Dado que el sistema tiene por lo menos un tipo de defecto, ¿cuál es la probabilidad de que tenga exactamente un tipo de defecto?

d. Dado que el sistema tiene los primeros dos tipos de defecto, ¿cuál es la probabilidad de que no tenga el tercer tipo de defecto?

49. La tabla adjunta proporciona información sobre el tipo de café seleccionado por alguien que compra una taza en un kiosco del aeropuerto en particular

	Pequeño	**Mediano**	**Grande**
Regular	14%	20%	26%
Descafeinado	20%	10%	10%

Considere la posibilidad de seleccionar al azar un comprador de café.

a. ¿Cuál es la probabilidad de que la persona adquiera una taza pequeña? ¿Una taza de café descafeinado?

b. Si nos enteramos de que la persona seleccionada compra una taza pequeña, ¿cuál es ahora la probabilidad de que él/ella escoja el café descafeinado y cómo interpreta esta probabilidad?

c. Si nos enteramos de que el individuo seleccionado compró un café descafeinado, ¿cuál es ahora la probabilidad de que un tamaño pequeño fue el escogido, y cómo se compara esto con la probabilidad incondicional correspondiente de (a)?

50. Una tienda de departamentos vende camisas sport en tres tallas (pequeña, mediana y grande), tres diseños (a cuadros, estampadas y a rayas) y dos largos de manga (larga y corta). Las tablas adjuntas dan las proporciones de camisas vendidas en las varias combinaciones de categoría.

Manga corta

	Diseño		
Talla	**A cuadros**	**Estampada**	**Rayas**
Ch	.04	.02	.05
M	.08	.07	.12
G	.03	.07	.08

Manga larga

	Diseño		
Talla	**A cuadros**	**Estampada**	**Rayas**
Ch	.03	.02	.03
M	.10	.05	.07
G	.04	.02	.08

a. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea una camisa mediana, estampada, de manga larga?
b. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea una camisa estampada, mediana?
c. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea de manga corta? ¿De manga larga?
d. ¿Cuál es la probabilidad de que la talla de la siguiente camisa vendida sea mediana? ¿Que la siguiente camisa vendida sea estampada?
e. Dado que la camisa que se acaba de vender era de manga corta a cuadros, ¿cuál es la probabilidad de que fuera mediana?
f. Dado que la camisa que se acaba de vender era mediana a cuadros, ¿cuál es la probabilidad de que fuera de manga corta? ¿De manga larga?

51. Una caja contiene seis pelotas rojas y cuatro verdes y una segunda caja contiene siete pelotas rojas y tres verdes. Se selecciona una pelota al azar de la primera caja y se la coloca en la segunda caja. Luego se selecciona al azar una pelota de la segunda caja y se la coloca en la primera caja.
a. ¿Cuál es la probabilidad de que se seleccione una pelota roja de la primera caja y de que se seleccione una pelota roja de la segunda caja?
b. Al final del proceso de selección, ¿cuál es la probabilidad de que los números de pelotas rojas y verdes que hay en la primera caja sean idénticos a los números iniciales?

52. Un sistema se compone de bombas idénticas, #1 y #2. Si una falla, el sistema seguirá operando. Sin embargo, debido al esfuerzo adicional, ahora es más probable que la bomba restante falle de lo que era originalmente. Es decir, $r = P(\#2 \text{ falla} \mid \#1 \text{ falla}) > P(\#2 \text{ falla}) = q$. Si por lo menos una bomba falla alrededor del final de su vida útil en 7% de todos los sistemas y ambas bombas fallan durante dicho periodo en sólo 1%, ¿cuál es la probabilidad de que la bomba #1 falle durante su vida útil?

53. Un taller repara componentes tanto de audio como de video. Sea A el evento en que el siguiente componente traído a reparación es un componente de audio, y sea B el evento en que el siguiente componente es un reproductor de discos compactos (así que el evento B está contenido en A). Suponga que $P(A) = .6$ y $P(B) = .05$. ¿Cuál es $P(B \mid A)$?

54. En el ejercicio 13, A_i = {proyecto otorgado i}, con $i = 1, 2, 3$. Use las probabilidades dadas allí para calcular las siguientes probabilidades y explique en palabras el significado de cada una.
a. $P(A_2 \mid A_1)$ **b.** $P(A_2 \cap A_3 \mid A_1)$
c. $P(A_2 \cup A_3 \mid A_1)$ **d.** $P(A_1 \cap A_2 \cap A_3 \mid A_1 \cup A_2 \cup A_3)$.

55. Las garrapatas de venados pueden ser portadoras de la enfermedad de Lyme o de la erliquiosis granulocítica humana (HGE, por sus siglas en inglés). Con base en un estudio reciente, suponga que 16% de todas las garrapatas en cierto lugar portan la enfermedad de Lyme, 10% portan HGE y 10% de las garrapatas que portan por lo menos una de estas enfermedades en realidad portan las dos. Si se determina que una garrapata seleccionada al azar ha sido portadora de HGE, ¿cuál es la probabilidad de que la garrapata seleccionada también porte la enfermedad de Lyme?

56. Para los eventos A y B con $P(B) > 0$, demuestre que $P(A \mid B) + P(A' \mid B) = 1$.

57. Si $P(B \mid A) > P(B)$, demuestre que $P(B' \mid A) < P(B')$. [*Sugerencia*: sume $P(B' \mid A)$ a ambos lados de la desigualdad dada y luego utilice el resultado del ejercicio 56.]

58. Demuestre que para tres eventos cualesquiera A, B y C con $P(C) > 0$, $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$.

59. En una gasolinera, 40% de los clientes utilizan gasolina regular (A_1), 35% usan gasolina plus (A_2) y 25% utilizan premium (A_3). De los clientes que utilizan gasolina regular, sólo 30% llenan sus tanques (evento B). De los clientes que utilizan plus, 60% llenan sus tanques, mientras que los que utilizan premium, 50% llenan sus tanques.
a. ¿Cuál es la probabilidad de que el siguiente cliente pida gasolina plus y llene el tanque ($A_2 \cap B$)?
b. ¿Cuál es la probabilidad de que el siguiente cliente llene el tanque?
c. Si el siguiente cliente llena el tanque, ¿cuál es la probabilidad que pida gasolina regular? ¿Plus? ¿Premium?

60. 70% de las aeronaves ligeras que desaparecen en vuelo en cierto país son posteriormente localizadas. De las aeronaves que son localizadas, 60% cuentan con un localizador de emergencia, mientras que 90% de las aeronaves no localizadas no cuentan con dicho localizador. Suponga que una aeronave ligera ha desaparecido.
a. Si tiene un localizador de emergencia, ¿cuál es la probabilidad de que no será localizada?
b. Si no tiene un localizador de emergencia, ¿cuál es la probabilidad de que será localizada?

61. Componentes de cierto tipo son enviados a un distribuidor en lotes de diez. Suponga que 50% de dichos lotes no contienen componentes defectuosos, 30% contienen un componente defectuoso y 20% contienen dos componentes defectuosos. Se seleccionan al azar dos componentes de un lote y se prueban.

¿Cuáles son las probabilidades asociadas con 0, 1 y 2 componentes defectuosos que están en el lote en cada una de las siguientes condiciones?

a. Ningún componente probado está defectuoso.

b. Uno de los dos componentes probados está defectuoso. [*Sugerencia*: trace un diagrama de árbol con tres ramas de primera generación correspondientes a los tres tipos diferentes de lotes.]

62. Una compañía que fabrica cámaras de video produce un modelo básico y un modelo de lujo. Durante el año pasado, 40% de las cámaras vendidas fueron del modelo básico. De aquellos que compraron el modelo básico, 30% adquirieron una garantía ampliada, en tanto que 50% de los que compraron el modelo de lujo también lo hicieron. Si se sabe que un comprador seleccionado al azar tiene una garantía ampliada, ¿qué tan probable es que él o ella tengan un modelo básico?

63. Para los clientes que compran un refrigerador en una tienda de aparatos domésticos, sea A el evento en que el refrigerador fue fabricado en EU, B el evento en que el refrigerador contaba con una máquina de hacer hielos y C el evento en que el cliente adquirió una garantía ampliada. Las probabilidades pertinentes son

$$P(A) = .75 \quad P(B|A) = .9 \quad P(B|A') = .8$$
$$P(C|A \cap B) = .8 \quad P(C|A \cap B') = .6$$
$$P(C|A' \cap B) = .7 \quad P(C|A' \cap B') = .3$$

a. Construya un diagrama de árbol compuesto de ramas de primera, segunda y tercera generaciones y anote el evento y la probabilidad apropiada junto a cada rama.

b. Calcule $P(A \cap B \cap C)$.

c. Calcule $P(B \cap C)$.

d. Calcule $P(C)$.

e. Calcule $P(A|B \cap C)$, la probabilidad de la compra de un refrigerador fabricado en EU dado que también se adquirieron una máquina de hacer hielos y una garantía ampliada.

64. El editor de comentarios de una cierta revista científica decide si la revisión de cualquier libro en particular debe ser corta (1–2 páginas), mediana (3–4 páginas), o larga (5–6 páginas). Los datos sobre estudios recientes indican que el 60% de ellas son cortas, el 30% son medianas y el otro 10% son largas. Los comentarios están presentados en Word o LaTeX. Para las revisiones cortas, el 80% son en Word, mientras que el 50% de las revisiones medianas se encuentran en Word y el 30% de las revisiones largas están en Word. Supongamos que se selecciona aleatoriamente una revisión reciente.

a. ¿Cuál es la probabilidad de que la revisión seleccionada se presentara en formato Word?

b. Si la revisión seleccionada se presentó en formato Word, ¿cuáles son las probabilidades posteriores de que sea corto, mediano o largo?

65. Un gran operador de complejos de tiempo compartido requiere que cualquier persona interesada en hacer una compra primero visite el sitio de interés. Los datos históricos indican que el 20% de todos los compradores potenciales seleccionaron un día de visita, el 50% elige una visita de una noche y el 30% opta por una visita de dos noches. Además, el 10% de los visitantes de un día en última instancia, hacen una compra, el 30% de los visitantes de una noche compran una unidad y el 20% de los visitantes de dos noches deciden comprar. Supongamos que un visitante es seleccionado al azar y se demuestra que ha realizado una compra. ¿Qué tan probable es que esta persona haya realizado una visita de día? ¿Una visita de una noche? ¿Una visita de dos noches?

66. Considere la siguiente información sobre vacacionistas (basada en parte en una encuesta reciente de Travelocity): 40% revisan su correo electrónico de trabajo, 30% utilizan un teléfono celular para permanecer en contacto con su trabajo, 25% trajeron una computadora portátil consigo, 23% revisan su correo electrónico de trabajo y utilizan un teléfono celular para permanecer en contacto, y 51% ni revisan su correo electrónico de trabajo ni utilizan un teléfono celular para permanecer en contacto ni trajeron consigo una computadora portátil. Además, 88 de cada 100 que traen una computadora portátil también revisan su correo electrónico de trabajo y 70 de cada 100 que utilizan un teléfono celular para permanecer en contacto también traen una computadora portátil.

a. ¿Cuál es la probabilidad de que un vacacionista seleccionado al azar que revisa su correo electrónico de trabajo también utilice un teléfono celular para permanecer en contacto?

b. ¿Cuál es la probabilidad de que alguien que trae una computadora portátil también utilice un teléfono celular para permanecer en contacto?

c. Si el vacacionista seleccionado al azar revisó su correo electrónico de trabajo y trajo una computadora portátil, ¿cuál es la probabilidad de que él o ella utilice un teléfono celular para permanecer en contacto?

67. Ha habido una gran controversia durante los últimos años con respecto a qué tipos de vigilancia son apropiados para impedir el terrorismo. Suponga que un sistema de vigilancia particular tiene 99% de probabilidades de identificar correctamente a un futuro terrorista y 99.9% de probabilidades de identificar correctamente a alguien que no es un futuro terrorista. Si existen 1000 futuros terroristas en una población de 300 millones y se selecciona al azar uno de estos 300 millones, es examinado por el sistema e identificado como futuro terrorista, ¿cuál es la probabilidad de que él o ella sea en realidad un futuro terrorista? ¿Le inquieta el valor de esta probabilidad sobre el uso del sistema de vigilancia? Explique.

68. Una amiga que vive en Los Ángeles hace viajes frecuentes de consultoría a Washington, D.C.; 50% del tiempo viaja en la línea aérea #1, 30% del tiempo en la aerolínea #2 y el 20% restante en la aerolínea #3. Los vuelos de la aerolínea #1 llegan demorados a D.C. 30% del tiempo y 10% del tiempo llegan demorados a L.A. Para la aerolínea #2, estos porcentajes son 25% y 20%, en tanto que para la aerolínea #3 los porcentajes son 40% y 25%. Si se sabe que en un viaje particular ella llegó demorada a exactamente uno de los dos destinos, ¿cuáles son las probabilidades posteriores de haber volado en las aerolíneas #1, #2 y #3? Suponga que la probabilidad de arribar con demora a L.A. no se ve afectada por lo que suceda en el vuelo a D.C. [*Sugerencia*: desde la punta de cada rama de primera generación en un diagrama de árbol, trace tres ramas de segunda generación identificadas, respectivamente, como 0 demorado, 1 demorado y 2 demorado.]

69. En el ejercicio 59, considere la siguiente información adicional sobre el uso de tarjetas de crédito:

70% de todos los clientes que utilizan gasolina regular y que llenan el tanque usan una tarjeta de crédito.

50% de todos los clientes que utilizan gasolina regular y que no llenan el tanque usan una tarjeta de crédito.
60% de todos los clientes que llenan el tanque con gasolina plus usan una tarjeta de crédito.
50% de todos los clientes que utilizan gasolina plus y que no llenan el tanque usan una tarjeta de crédito.
50% de todos los clientes que utilizan gasolina premium y que llenan el tanque usan una tarjeta de crédito.
40% de todos los clientes que utilizan gasolina premium y que no llenan el tanque usan una tarjeta de crédito.

Calcule la probabilidad de cada uno de los siguientes eventos para el siguiente cliente que llegue (un diagrama de árbol podría ayudar).

a. {plus, tanque lleno y tarjeta de crédito}
b. {premium, tanque no lleno y tarjeta de crédito}
c. {premium y tarjeta de crédito}
d. {tanque lleno y tarjeta de crédito}
e. {tarjeta de crédito}
f. Si el siguiente cliente utiliza una tarjeta de crédito, ¿cuál es la probabilidad de que pida premium?

2.5 Independencia

La definición de probabilidad condicional permite revisar la probabilidad $P(A)$ originalmente asignada a A cuando después se informa que otro evento B ha ocurrido; la nueva probabilidad de A es $P(A|B)$. En los ejemplos, con frecuencia fue el caso de que $P(A|B)$ difería de la probabilidad no condicional $P(A)$, lo que indica que la información "B ha ocurrido" cambia la probabilidad de que ocurra A. A menudo la probabilidad de que ocurra o haya ocurrido A no se ve afectada por el conocimiento de que B ha ocurrido, así que $P(A|B) = P(A)$. Es entonces natural considerar a A y B como eventos independientes, es decir que la ocurrencia o no ocurrencia de un evento no afecta la probabilidad de que el otro ocurra.

DEFINICIÓN

Los eventos A y B son **independientes** si $P(A|B) = P(A)$ y son **dependientes** en caso contrario.

La definición de independencia podría parecer "no simétrica" porque no demanda también que $P(B|A) = P(B)$. Sin embargo, utilizando la definición de probabilidad condicional y la regla de multiplicación,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \tag{2.7}$$

El lado derecho de la ecuación (2.7) es $P(B)$ si y sólo si $P(A|B) = P(A)$ (independencia), así que la igualdad en la definición implica la otra igualdad (y viceversa). También es fácil demostrar que si A y B son independientes, entonces también lo son los pares de eventos: (1) A' y B, (2) A y B' y (3) A' y B'.

Ejemplo 2.32 Considere una gasolinera con seis bombas numeradas 1, 2, . . . , 6, y sea E_i el evento simple en que un cliente seleccionado al azar utiliza la bomba i ($i = 1, \ldots, 6$). Suponga que

$$P(E_1) = P(E_6) = .10, \quad P(E_2) = P(E_5) = .15, \quad P(E_3) = P(E_4) = .25$$

Defina los eventos A, B, C como

$$A = \{2, 4, 6\}, B = \{1, 2, 3\}, C = \{2, 3, 4, 5\}.$$

Luego se tiene $P(A) = .50$, $P(A|B) = .30$ y $P(A|C) = .50$. Es decir, los eventos A y B son dependientes, en tanto que los eventos A y C son independientes. Intuitivamente, A y C son independientes porque la división de probabilidad relativa entre las bombas pares e impares es la misma entre las bombas 2, 3, 4, 5, como lo es entre todas las seis bombas. ■

Ejemplo 2.33 Sean A y B dos eventos mutuamente exclusivos cualesquiera con $P(A) > 0$. Por ejemplo, para un automóvil seleccionado al azar, sea A = {el carro es de cuatro cilindros} y B = {el carro es de seis cilindros}. Como los eventos son mutuamente exclusivos, si B ocurre, entonces A quizás puede no haber ocurrido, así que $P(A|B) = 0 \neq P(A)$. El mensaje aquí es que *si dos eventos son mutuamente exclusivos, no pueden ser independientes*.

Cuando A y B son mutuamente exclusivos, la información de que A ocurrió dice algo sobre B (no puede haber ocurrido), así que se impide la independencia. ■

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

Con frecuencia la naturaleza de un experimento sugiere que dos eventos A y B deben ser supuestos independientes. Este es el caso, por ejemplo, si un fabricante recibe una tarjeta de circuito de cada uno de dos proveedores diferentes, cada tarjeta se somete a prueba al llegar y $A = \{$la primera está defectuosa$\}$ y $B = \{$la segunda está defectuosa$\}$. Si $P(A) = .1$, también deberá ser el caso de que $P(A \mid B) = .1$; sabiendo la condición de la segunda tarjeta no informa sobre la condición de la primera. La probabilidad de que ambos eventos ocurran se calcula fácilmente a partir de la probabilidad individual de los eventos cuando éstos son independientes.

PROPOSICIÓN

> A y B son independientes si y sólo si
>
> $$P(A \cap B) = P(A) \cdot P(B) \qquad (2.8)$$

La verificación de esta regla de multiplicación es como sigue:

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B) \qquad (2.9)$$

donde la segunda igualdad en la ecuación (2.9) es válida si y sólo si A y B son independientes. Debido a la equivalencia de independencia y la ecuación (2.8), la segunda puede ser utilizada como definición de independencia.

Ejemplo 2.34 Se sabe que 30% de las lavadoras de cierta compañía requieren servicio mientras se encuentran dentro de garantía, en tanto que sólo 10% de sus secadoras necesitan dicho servicio. Si alguien adquiere tanto una lavadora como una secadora fabricadas por esta compañía, ¿cuál es la probabilidad de que ambas máquinas requieran servicio de garantía?

Sea A el evento en que la lavadora necesita servicio mientras se encuentra dentro de garantía y defina B de forma análoga para la secadora. Entonces $P(A) = .30$ y $P(B) = .10$. Suponiendo que las dos máquinas funcionan independientemente una de otra, la probabilidad deseada es

$$P(A \cap B) = P(A) \cdot P(B) = (.30)(.10) = .03$$ ■

Es fácil demostrar que A y B son independientes si y sólo si A' y B son independientes, A y B' son independientes y A' y B' son independientes. Por lo tanto, en el ejemplo 2.34, la probabilidad de que ninguna máquina necesite servicio es

$$P(A' \cap B') = P(A') \cdot P(B') = (.70)(.90) = .63$$

Ejemplo 2.35 Cada día, de lunes a viernes, un lote de componentes enviado por un primer proveedor arriba a una instalación de inspección. Dos días a la semana, también arriba un lote de un segundo proveedor. Ochenta por ciento de todos los lotes del proveedor 1 son inspeccionados y 90% de los del proveedor 2 también lo son. ¿Cuál es la probabilidad de que, en un día seleccionado al azar, dos lotes sean inspeccionados? Esta pregunta se responderá suponiendo que en los días en que se inspeccionan dos lotes, si el primer lote pasa es independiente de si el segundo también lo hace. La figura 2.13 muestra la información relevante.

$$\begin{aligned} P(\text{dos pasan}) &= P(\text{dos recibidos} \cap \text{ambos pasan}) \\ &= P(\text{ambos pasan} \mid \text{dos recibidos}) \cdot P(\text{dos recibidos}) \\ &= [(.8)(.9)](.4) = .288 \end{aligned}$$ ■

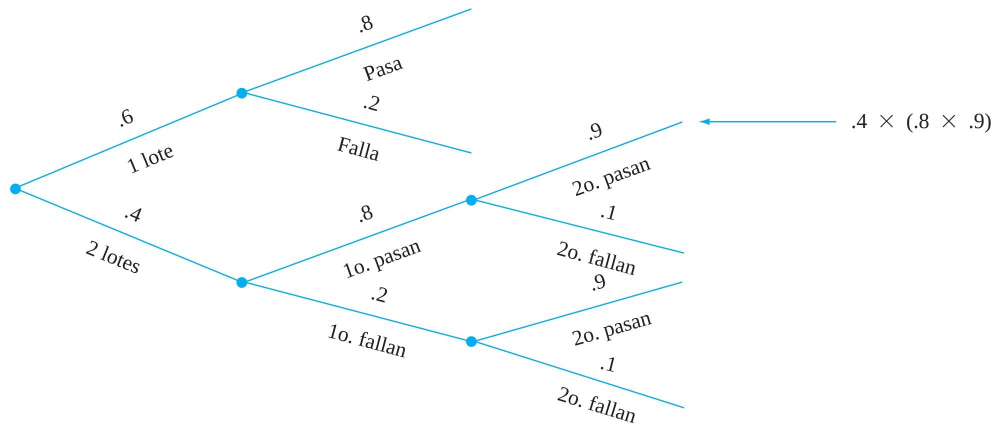


Figura 2.13 Diagrama de árbol para el ejemplo 2.35

Independencia de más de dos eventos

La noción de independencia de dos eventos puede ser ampliada a conjuntos de más de dos eventos. Aunque es posible ampliar la definición para dos eventos independientes trabajando en función de probabilidades condicionales y no condicionales, es más directo y menos tedioso seguir las líneas de la última proposición.

DEFINICIÓN

Los eventos $A_1, \ldots, A_n$ son **mutuamente independientes** si por cada k ($k = 2, 3, \ldots, n$) y cada subconjunto de índices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k})$$

Parafraseando la definición, los eventos son mutuamente independientes si la probabilidad de la intersección de cualquier subconjunto de los n eventos es igual al producto de las probabilidades individuales. Al utilizar la propiedad de multiplicación para más de dos eventos independientes, es legítimo reemplazar una o más de las A_i por sus complementos (p. ej., si A_1, A_2 y A_3 son eventos independientes, también lo son A_1', A_2' y A_3'). Como fue el caso con dos eventos, con frecuencia se especifica al principio de un problema la independencia de ciertos eventos. La probabilidad de una intersección puede entonces ser calculada vía multiplicación.

Ejemplo 2.36 El artículo "Reliability Evaluation of Solar Photovoltaic Arrays" (*Solar Energy*, 2002: 129–141) presenta varias configuraciones de redes fotovoltaicas solares compuestas de celdas solares de silicio cristalino. Considérese primero el sistema ilustrado en la figura 2.14(a).

Existen dos subsistemas conectados en paralelo, y cada uno contiene tres celdas. Para que el sistema funcione, por lo menos uno de los dos subsistemas en paralelo debe funcio-

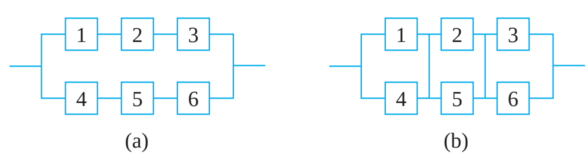


Figura 2.14 Configuraciones de sistema para el ejemplo 2.36: (a) serie-paralelo; (b) total-transversal-ligada

nar. Dentro de cada subsistema, las tres celdas están conectadas en serie, así que un subsistema funcionará sólo si todas sus celdas funcionan. Considere un valor de duración particular t_0 y suponga que desea determinar la probabilidad de que la duración del sistema exceda de t_0. Sea A_i el evento en que la duración de la celda i excede de t_0 ($i = 1, 2, \ldots, 6$). Se supone que las A_i' son eventos independientes (ya sea que cualquier celda particular que dure más de t_0 horas no tenga ningún efecto en sí o no cualquier otra celda lo hace) y que $P(A_i) = .9$ por cada i, puesto que las celdas son idénticas. Entonces

$$\begin{aligned} P(\text{la duración del sistema excede de } t_0) &= P[(A_1 \cap A_2 \cap A_3) \cup (A_4 \cap A_5 \cap A_6)] \\ &= P(A_1 \cap A_2 \cap A_3) + P(A_4 \cap A_5 \cap A_6) \\ &\quad - P[(A_1 \cap A_2 \cap A_3) \cap (A_4 \cap A_5 \cap A_6)] \\ &= (.9)(.9)(.9) + (.9)(.9)(.9) - (.9)(.9)(.9)(.9)(.9)(.9) = .927 \end{aligned}$$

Alternativamente,

$$\begin{aligned} P(\text{la duración del sistema excede de } t_0) &= 1 - P(\text{ambas duraciones del subsistema son } \le t_0) \\ &= 1 - [P(\text{la duración del subsistema es } \le t_0)]^2 \\ &= 1 - [1 - P(\text{la duración del subsistema es } > t_0)]^2 \\ &= 1 - [1 - (.9)^3]^2 = .927 \end{aligned}$$

Considérese a continuación el sistema vinculado en cruz mostrado en la figura 2.14(b), obtenido a partir de la red conectada en serie-paralelo mediante la conexión de enlaces a través de cada columna de uniones. Ahora el sistema falla en cuanto toda una columna falla, y la duración del sistema excede de t_0 sólo si la duración de cada columna lo hace. Para esta configuración,

$$\begin{aligned} P(\text{la duración del sistema es de por lo menos } t_0) &= [P(\text{la duración de la columna excede de } t_0)]^3 \\ &= [1 - P(\text{duración de la columna es } \le t_0)]^3 \\ &= [1 - P(\text{la duración de ambas celdas en una columna es } \le t_0)]^3 \\ &= [1 - (1 - .9)^2]^3 = .970 \end{aligned}$$

EJERCICIOS Sección 2.5 (70–89)

70. Reconsidere el escenario de la tarjeta de crédito del ejercicio 47 (sección 2.4) y demuestre que A y B son dependientes utilizando primero la definición de independencia y luego verificando que la propiedad de multiplicación no prevalece.

71. Una compañía de exploración petrolera en la actualidad tiene dos proyectos activos, uno en Asia y el otro en Europa. Sea A el evento en que el proyecto asiático tiene éxito y B el evento en que el proyecto europeo tiene éxito. Suponga que A y B son eventos independientes con $P(A) = .4$ y $P(B) = .7$.

a. Si el proyecto asiático no tiene éxito, ¿cuál es la probabilidad de que el europeo tampoco tenga éxito? Explique su razonamiento.

b. ¿Cuál es la probabilidad de que por lo menos uno de los dos proyectos tenga éxito?

c. Dado que por lo menos uno de los dos proyectos tiene éxito, ¿cuál es la probabilidad de que sólo el proyecto asiático tenga éxito?

72. En el ejercicio 13, ¿es cualquier A_i independiente de cualquier otro A_j? Responda utilizando la propiedad de multiplicación para eventos independientes.

73. Si A y B son eventos independientes, demuestre que A' y B también son independientes. [*Sugerencia*: primero establezca una relación entre $P(A' \cap B)$, $P(B)$ y $P(A \cap B)$.]

74. Suponga que las proporciones de fenotipos sanguíneos en una población son las siguientes:

A	B	AB	O
.40	.11	.04	.45

Suponiendo que los fenotipos de dos individuos seleccionados al azar son independientes uno de otro, ¿cuál es la probabilidad de que ambos fenotipos sean O? ¿Cuál es la probabilidad de que los fenotipos de dos individuos seleccionados al azar coincidan?

75. Una de las suposiciones que sustentan la teoría de las gráficas de control (véase el capítulo 16) es que los puntos graficados sucesivos son independientes entre sí. Cada punto puede señalar que un proceso de producción está funcionando correctamente o que existe algún funcionamiento defectuoso. Aun cuando un proceso esté funcionando de manera correcta, existe una pequeña probabilidad de que un punto particular señalará un problema con el proceso. Suponga que esta probabilidad es de .05. ¿Cuál es la probabilidad de que por lo menos uno de 10 puntos sucesivos indique un problema cuando de hecho el proceso está operando correctamente? Responda esta pregunta para 25 puntos sucesivos.

76. En octubre de 1994 se descubrió un defecto en un determinado chip Pentium instalado en las computadoras que podía dar lugar a una respuesta equivocada al realizar una división. El fabricante sostuvo inicialmente que la posibilidad de que cualquier división particular fuera incorrecta era sólo 1 de cada 9 mil millones, así que tomaría miles de años antes de que un usuario típico detectara un error. Sin embargo, los estadísticos no son usuarios típicos, algunas técnicas estadísticas modernas son tan computacionalmente intensivas que mil millones de divisiones en un corto periodo de tiempo no están fuera del reino de la posibilidad. Suponiendo que la cifra de 1 en 9 mil millones es correcta y que los resultados de las distintas divisiones son independientes el uno del otro, ¿cuál es la probabilidad de que al menos haya un error en mil millones de divisiones con este chip?

77. La costura de un avión requiere 25 remaches. La costura tendrá que ser retrabajada si alguno de los remaches está defectuoso. Suponga que los remaches están defectuosos independientemente uno de otro, cada uno con la misma probabilidad.

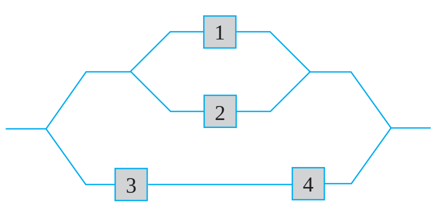
a. Si 20% de todas las costuras tienen que ser retrabajadas, ¿cuál es la probabilidad de que un remache esté defectuoso?

b. ¿Qué tan pequeña deberá ser la probabilidad de un remache defectuoso para garantizar que sólo 10% de las costuras tengan que ser retrabajadas?

78. Una caldera tiene cinco válvulas de alivio idénticas. La probabilidad de que cualquier válvula particular se abra en un momento de demanda es de .95. Suponiendo que operan independientemente, calcule P(por lo menos una válvula se abre) y P(por lo menos una válvula no se abre).

79. Dos bombas conectadas en paralelo fallan independientemente una de otra en cualquier día dado. La probabilidad de que falle sólo la bomba más vieja es de .10 y la probabilidad de que sólo la bomba más nueva falle es de .05. ¿Cuál es la probabilidad de que el sistema de bombeo falle en cualquier día dado (lo que sucede si ambas bombas fallan)?

80. Considere el sistema de componentes conectados como en la figura adjunta. Los componentes 1 y 2 están conectados en paralelo, de modo que el subsistema trabaja si y sólo si 1 o 2 trabaja; como 3 y 4 están conectados en serie, ¿el subsistema trabaja si y sólo si 3 y 4 trabajan? Si los componentes funcionan independientemente uno de otro y P(el componente trabaja) $= .9$, calcule P(el sistema trabaja).



81. Remítase otra vez al sistema en serie-paralelo introducido en el ejemplo 2.35 y suponga que existen sólo dos celdas en lugar de tres en cada subsistema en paralelo [en la figura 2.14(a), elimine las celdas 3 y 6 y renumere las celdas 4 y 5 como 3 y 4]. Utilizando $P(A_i) = .9$, es fácil ver que la probabilidad de que la duración del sistema exceda de t_0 es de .9639. ¿A qué valor tendría que cambiar .9 para incrementar la confiabilidad y duración del sistema de .9639 a .99? [*Sugerencia*: sea $P(A_i) = p$, exprese la confiabilidad del sistema en función de p, luego haga $x = p^2$.]

82. Considere lanzar en forma independiente dos dados imparciales, uno rojo y otro verde. Sea A el evento en que el dado rojo muestra 3 puntos, B el evento en que el dado verde muestra 4 puntos y C el evento en que el número total de puntos que muestran los dos dados es 7. ¿Son estos eventos independientes por pares (es decir, ¿son A y B eventos independientes, son A y C independientes y son B y C independientes? ¿Son los tres eventos mutuamente independientes?

83. Los componentes enviados a un distribuidor son revisados en cuanto a defectos por dos inspectores diferentes (cada componente es revisado por ambos inspectores). El primero detecta 90% de todos los defectuosos que están presentes y el segundo hace lo mismo. Por lo menos un inspector no detecta un defecto en 20% de todos los componentes defectuosos. ¿Cuál es la probabilidad de que ocurra lo siguiente?

a. ¿Un componente defectuoso será detectado sólo por el primer inspector? ¿Por exactamente uno de los dos inspectores?

b. ¿Los tres componentes defectuosos en un lote no son detectados por ambos inspectores (suponiendo que las inspecciones de los diferentes componentes son independientes unas de otras)?

84. Setenta por ciento de todos los vehículos examinados en un centro de verificación de emisiones pasan la inspección. Suponiendo que vehículos sucesivos pasan o fallan independientemente uno de otro, calcule las siguientes probabilidades:

a. P(los tres vehículos siguientes inspeccionados pasan)

b. P(por lo menos uno de los tres vehículos siguientes pasa)

c. P(exactamente uno de los tres vehículos siguientes pasa)

d. P(cuando mucho uno de los tres vehículos siguientes inspeccionados pasa)

e. Dado que por lo menos uno de los tres vehículos siguientes pasa la inspección, ¿cuál es la probabilidad de que los tres pasen (una probabilidad condicional)?

85. Un inspector de control de calidad verifica artículos recién producidos en busca de fallas. El inspector examina un artículo en busca de fallas en una serie de observaciones independientes, cada una de duración fija. Dado que en realidad está presente una imperfección, sea p la probabilidad de que la imperfección sea detectada durante cualquier observación (este modelo se discute en "Human Performance in Sampling Inspection", *Human Factors*, 1979: 99–105).

a. Suponiendo que un artículo tiene una imperfección, ¿cuál es la probabilidad de que sea detectada al final de la segunda observación (una vez que una imperfección ha sido detectada, la secuencia de observaciones termina)?

b. Dé una expresión para la probabilidad de que una imperfección será detectada al final de la n-ésima observación.

c. Si cuando en tres observaciones no ha sido detectada una imperfección, el artículo es aprobado, ¿cuál es la probabilidad de que un artículo imperfecto pase la inspección?

d. Suponga que 10% de todos los artículos contienen una imperfección [P(artículo seleccionado al azar muestra una imperfección) $= .1$]. Con la suposición del inciso (c), ¿cuál es la probabilidad de que un artículo seleccionado al azar pase la inspección (pasará automáticamente si no muestra una imperfección, pero también podría pasar si muestra una imperfección)?

e. Dado que un artículo ha pasado la inspección (ninguna imperfección en tres observaciones), ¿cuál es la probabilidad de que sí tenga una imperfección? Calcule para $p = .5$.

86. a. Una compañía maderera acaba de recibir un lote de 10,000 tablas de 2×4. Suponga que 20% de estas tablas (2,000) en realidad están demasiado tiernas o verdes para ser utilizadas en construcción de primera calidad. Se eligen dos tablas al azar, una después de la otra. *Sea* $A = \{$la primera tabla está verde$\}$ y $B = \{$la segunda tabla está verde$\}$. Calcule $P(A)$, $P(B)$ y $P(A \cap B)$ (un diagrama de árbol podría ayudar). ¿Son A y B independientes?

b. Con A y B independientes y $P(A) = P(B) = .2$, ¿cuál es $P(A \cap B)$? ¿Cuánta diferencia existe entre esta respuesta y $P(A \cap B)$ en el inciso (a)? Para propósitos de cálculo $P(A \cap B)$, ¿se puede suponer que A y B del inciso (a) son independientes para obtener en esencia la probabilidad correcta?

c. Suponga que un lote consta de 10 tablas, de las cuales dos están verdes. ¿Produce ahora la suposición de independencia aproximadamente la respuesta correcta para $P(A \cap B)$? ¿Cuál es la diferencia crítica entre la situación en este caso y la del inciso (a)? ¿Cuándo piensa que una suposición de independencia sería válida al obtener una respuesta aproximadamente correcta para $P(A \cap B)$?

87. Considere la posibilidad de seleccionar al azar una sola persona y que ésta prueba tres vehículos diferentes. Defina los eventos A_1, A_2 y A_3 por

A_1 = como el vehículo #1 $\quad$ A_2 = como el vehículo #2

A_3 = como el vehículo #3

Suponga que $P(A_1) = .55$, $P(A_2) = .65$, $P(A_3) = .70$, $P(A_1 \cup A_2) = .80$, $P(A_2 \cap A_3) = .40$ y $P(A_1 \cup A_2 \cup A_3) = .88$.

a. ¿Cuál es la probabilidad de que a la persona le guste tanto el vehículo # 1 como el vehículo # 2?

b. Determinar e interpretar $P(A_2|A_3)$.

c. ¿Son eventos independientes A_2 y A_3? Responda en dos maneras diferentes.

d. Si usted se entera de que a la persona no le gusta el vehículo # 1, ¿cuál es ahora la probabilidad de que a él/ella le guste por lo menos uno de los otros dos vehículos?

88. El profesor Stan der Deviation puede tomar una de dos rutas en el trayecto del trabajo a su casa. En la primera ruta, hay cuatro cruces de ferrocarril. La probabilidad de que sea detenido por un tren en cualquiera de los cruces es .1 y los trenes operan independientemente en los cuatro cruces. La otra ruta es más larga pero sólo hay dos cruces, independientes uno de otro, con la misma posibilidad de que sea detenido por un tren al igual que en la primera ruta. En un día particular, el profesor Deviation tiene una reunión programada en casa durante cierto tiempo. ¿Cualquier ruta que tome, calcula que llegará tarde si es detenido por los trenes en por lo menos la mitad de los cruces encontrados.

a. ¿Cuál ruta deberá tomar para reducir al mínimo la probabilidad de llegar tarde a la reunión?

b. Si lanza al aire una moneda imparcial para decidir qué ruta tomar y está retrasado, ¿cuál es la probabilidad de que tome la ruta de los cuatro cruces?

89. Suponga que se colocan etiquetas idénticas en las dos orejas de un zorro. El zorro es dejado en libertad durante un lapso de tiempo. Considere los dos eventos $C_1 = \{$se pierde la etiqueta de la oreja izquierda$\}$ y $C_2 = \{$se pierde la etiqueta de la oreja derecha$\}$. Sea $\pi = P(C_1) = P(C_2)$ y suponga que C_1 y C_2 son eventos independientes. Deduzca una expresión (que implique π) para la probabilidad de que exactamente una etiqueta se pierda dado que cuando mucho una se pierde ("Ear Tag Loss in Red Foxes", *J. Wildlife Mgmt*., 1976: 164–167). [*Sugerencia*: trace un diagrama de árbol en el cual las dos ramas iniciales se refieren a si la etiqueta de la oreja izquierda se pierde.]

EJERCICIOS SUPLEMENTARIOS (90–114)

90. Una pequeña compañía manufacturera va a echar a andar un turno de noche. Hay 20 mecánicos empleados por la compañía.

a. Si una cuadrilla nocturna se compone de 3 mecánicos, ¿cuántas cuadrillas diferentes son posibles?

b. Si los mecánicos están clasificados 1, 2, . . . , 20 en orden de competencia, ¿cuántas de estas cuadrillas no incluirían al mejor mecánico?

c. ¿Cuántas de las cuadrillas tendrían por lo menos 1 de los 10 mejores mecánicos?

d. Si se selecciona al azar una de estas cuadrillas para que trabajen una noche particular, ¿cuál es la probabilidad de que el mejor mecánico no trabaje esa noche?

91. Una fábrica utiliza tres líneas de producción para fabricar latas de cierto tipo. La tabla adjunta da porcentajes de latas que no cumplen con las especificaciones, clasificadas por tipo de incumplimiento de las especificaciones, para cada una de las tres líneas durante un lapso de tiempo particular.

	Línea 1	Línea 2	Línea 3
Manchas	15	12	20
Grietas	50	44	40
Problemas con la argolla	21	28	24
Defecto superficial	10	8	15
Otro	4	8	2

Durante este periodo, la línea 1 produjo 500 latas fuera de especificación, la 2 produjo 400 latas como ésas y la 3 fue responsable de 600 latas fuera de especificación. Suponga que se selecciona al azar una de estas 1500 latas.

a. ¿Cuál es la probabilidad de que la lata venga de la línea 1? ¿Cuál es la probabilidad de que la razón del incumplimiento de la especificación sea una grieta?

b. Si la lata seleccionada provino de la línea 1, ¿cuál es la probabilidad de que tenga una mancha?

c. Dado que la lata seleccionada mostró un defecto superficial, ¿cuál es la probabilidad de que provenga de la línea 1?

92. Un empleado de la oficina de inscripciones en una universidad en este momento tiene diez formas en su escritorio en espera de ser procesadas. Seis de éstas son peticiones de baja y las otras cuatro son solicitudes de sustitución de curso.

a. Si selecciona al azar seis de estas formas para dárselas a un subordinado, ¿cuál es la probabilidad de que sólo uno de los dos tipos de formas permanezca en su escritorio?

b. Suponga que tiene tiempo para procesar sólo cuatro de estas formas antes de salir del trabajo. Si estas cuatro se seleccionan al azar una por una, ¿cuál es la probabilidad de que cada forma subsiguiente sea de un tipo diferente de su predecesora?

93. Un satélite está programado para ser lanzado desde Cabo Cañaveral, en Florida, y otro lanzamiento está programado para la Base de la Fuerza Aérea Vandenberg en California. Sea A el evento en que el lanzamiento en Vandenberg se hace a la hora programada y B el evento en que el lanzamiento en Cabo Cañaveral se hace a la hora programada. Si A y B son eventos independientes con $P(A) > P(B)$, $P(A \cup B) = .626$ y $P(A \cap B) = .144$, determine los valores de $P(A)$ y $P(B)$.

94. Un transmisor envía un mensaje utilizando un código binario, esto es, una secuencia de ceros y unos. Cada bit transmitido (0 o 1) debe pasar a través de tres relevadores para llegar al receptor. En cada relevador, la probabilidad es .20 de que el bit enviado será diferente del bit recibido (una inversión). Suponga que los relevadores operan independientemente uno de otro.

transmisor $\rightarrow$ relevador 1 $\rightarrow$ relevador 2 $\rightarrow$ relevador 3 $\rightarrow$ receptor

a. Si el transmisor envía un 1, ¿cuál es la probabilidad de que los tres relevadores envíen un 1?

b. Si el transmisor envía un 1, ¿cuál es la probabilidad de que el receptor reciba un 1? [*Sugerencia*: los ocho resultados experimentales pueden ser mostrados en un diagrama de árbol con tres ramas de generación, una por cada relevador.]

c. Suponga que 70% de todos los bits enviados por el transmisor son unos. Si el receptor recibe un 1, ¿cuál es la probabilidad de que un 1 haya sido enviado?

95. El individuo A tiene un círculo de cinco amigos cercanos (B, C, D, E y F). A escuchó cierto rumor originado fuera del círculo e invitó a sus cinco amigos a una fiesta para contarles el rumor. Para empezar, A escoge a uno de los cinco al azar y se lo cuenta. Dicho individuo escoge entonces al azar a uno de los cuatro individuos restantes y repite el rumor. Después, de aquellos que ya oyeron el rumor uno se lo cuenta a otro nuevo individuo y así hasta que todos oyen el rumor.

a. ¿Cuál es la probabilidad de que el rumor se repita en el orden B, C, D, E y F?

b. ¿Cuál es la probabilidad de que F sea la tercera persona en la reunión a la que se le contará el rumor?

c. ¿Cuál es la probabilidad de que F sea la última persona en oír el rumor?

d. Si en cada etapa la persona que en ese momento “tiene” el rumor no sabe quien ya lo ha escuchado y selecciona al siguiente destinatario aleatoriamente de entre cinco individuos posibles, ¿cuál es la probabilidad de que F no haya escuchado todavía el rumor después de haber sido dicho 10 veces en la fiesta?

96. De acuerdo con el artículo “Optimization of Distribution Parameters for Estimating Probability of Crack Detection” (*J. of Aircraft*, 2009: 2090–2097), la siguiente ecuación de “Palmberg” se usa comúnmente para determinar la probabilidad $P_d(c)$ de la detección de una grieta de tamaño c en la estructura de la aeronave:

$$P_d(c) = \frac{(c/c^*)^\beta}{1 + (c/c^*)^\beta}$$

donde c^* es el tamaño de la grieta que corresponde a una probabilidad de detección de .5 (y por tanto es una evaluación de la calidad del proceso de inspección).

a. Compruebe que $P_d(c^*) = .5$

b. ¿Qué es $P_d(2c^*)$ cuando $\beta = 4$?

c. Supongamos que un inspector revisa dos paneles diferentes, uno con un tamaño de grieta de c^* y otro con un tamaño de grieta de $2c^*$. Una vez más, suponiendo $\beta = 4$ y también que los resultados de las dos inspecciones son independientes el uno del otro, ¿cuál es la probabilidad de que exactamente una de las dos grietas se detecte?

d. ¿Qué le sucede a $P_d(c)$ cuando $\beta \rightarrow \infty$?

97. Un ingeniero químico está interesado en determinar si cierta impureza está presente en un producto. Un experimento tiene una probabilidad de .80 de detectarla si está presente. La probabilidad de no detectarla si está ausente es de .90. Las probabilidades previas de que la impureza esté presente o ausente son de .40 y .60, respectivamente. Tres experimentos distintos producen sólo dos detecciones. ¿Cuál es la probabilidad posterior de que la impureza esté presente?

98. A cada concursante en un programa de preguntas se le pide que especifique una de seis posibles categorías de entre las cuales se le hará una pregunta. Suponga P(el concursante escoge la categoría i) $= \frac{1}{6}$ y concursantes sucesivos escogen sus categorías independientemente uno del otro. Si participan tres concursantes en cada programa y los tres en un programa particular seleccionan diferentes categorías, ¿cuál es la probabilidad de que exactamente uno seleccione la categoría 1?

99. Los sujetadores roscados utilizados en la fabricación de aviones son levemente doblados para que queden bien apretados y no se aflojen durante vibraciones. Suponga que 95% de todos los sujetadores pasan una inspección inicial. De 5% que fallan, 20% están tan seriamente defectuosos que deben ser desechados. Los sujetadores restantes son enviados a una operación de redoblado, donde 40% no pueden ser recuperados y son desechados. El otro 60% de estos sujetadores son corregidos por el proceso de redoblado y posteriormente pasan la inspección.

a. ¿Cuál es la probabilidad de que un sujetador que acaba de llegar seleccionado al azar pase la inspección inicialmente o después del redoblado?

b. Dado que un sujetador pasó la inspección, ¿cuál es la probabilidad de que apruebe la inspección inicial y de que no necesite redoblado?

100. Un porcentaje de todos los individuos en una población son portadores de una enfermedad particular. Una prueba de diagnóstico para esta enfermedad tiene una tasa de detección de 90% para portadores y de 5% para no portadores. Suponga que la prueba se aplica independientemente a dos muestras de sangre diferentes del mismo individuo seleccionado al azar.

a. ¿Cuál es la probabilidad de que ambas pruebas den el mismo resultado?

b. Si ambas pruebas son positivas, ¿cuál es la probabilidad de que el individuo seleccionado sea un portador?

101. Un sistema consta de dos componentes. La probabilidad de que el segundo componente funcione de manera satisfactoria durante su duración de diseño es de .9, la probabilidad de que por lo menos uno de los dos componentes lo haga es de .96 y la probabilidad de que ambos componentes lo hagan es de .75. Dado que el primer componente funciona de manera satisfactoria durante toda su duración de diseño, ¿cuál es la probabilidad de que el segundo también lo haga?

102. Cierta compañía envía 40% de sus paquetes de correspondencia nocturna vía un servicio de correo exprés E_1. De estos paquetes, 2% llegan después del tiempo de entrega garantizado (sea L el evento "entrega demorada"). Si se selecciona al azar un registro de correspondencia nocturna del archivo de la compañía, ¿cuál es la probabilidad de que el paquete se fue vía E_1 y llegó demorado?

103. Remítase al ejercicio 102. Suponga que 50% de los paquetes nocturnos se envían vía el servicio de correo exprés E_2 y el 10% restante se envía vía E_3. De los paquetes enviados vía E_2, sólo 1% llegaron demorados, en tanto que 5% de los paquetes manejados por E_3 llegaron demorados.

a. ¿Cuál es la probabilidad de que un paquete seleccionado al azar llegue demorado?

b. Si un paquete seleccionado al azar llegó a tiempo, ¿cuál es la probabilidad de que no haya sido enviado vía E_1?

104. Una compañía utiliza tres líneas de ensamble diferentes: A_1, A_2 y A_3 para fabricar un componente particular. De los fabricados por la línea A_1, 5% tienen que ser retrabajados para corregir un defecto, mientras que 8% de los componentes de A_2 tienen que ser retrabajados y 10% de los componentes de A_3 tienen que ser retrabajados. Suponga que 50% de todos los componentes los produce la línea A_1, 30% la línea A_2 y 20% la línea A_3. Si un componente seleccionado al azar tiene que ser retrabajado, ¿cuál es la probabilidad de que provenga de la línea A_1? ¿De la línea A_2? ¿De la línea A_3?

105. Desechando la posibilidad de cumplir años el 29 de febrero, suponga que es igualmente probable que un individuo seleccionado al azar haya nacido en cualquiera de los demás 365 días.

a. Si se seleccionan al azar diez personas, ¿cuál es la probabilidad de que tengan diferentes cumpleaños? ¿De que por lo menos dos tengan el mismo cumpleaños?

b. Si k reemplaza a diez en el inciso (a), ¿cuál es la k más pequeña para la cual existe por lo menos una probabilidad de 50-50 de que dos o más personas tengan el mismo cumpleaños?

c. Si se seleccionan diez personas al azar, ¿cuál es la probabilidad de que por lo menos dos tengan el mismo cumpleaños o por lo menos dos tengan los mismos tres últimos dígitos de sus números del Seguro Social? [*Nota*: el artículo "Methods for Studying Coincidences" (F. Mosteller y P. Diaconis, *J. Amer. Stat. Assoc.*, 1989: 853–861) discute problemas de este tipo.]

106. Un método utilizado para distinguir entre rocas graníticas (G) y basálticas (B) es examinar una parte del espectro infrarrojo de la energía solar reflejada por la superficie de la roca. Sean R_1, R_2 y R_3 intensidades espectrales medidas a tres longitudes de onda diferentes; en general, para granito $R_1 < R_2 < R_3$, en tanto que para basalto $R_3 < R_1 < R_2$. Cuando se hacen mediciones a distancia (mediante un avión), varios ordenamientos de R_i pueden presentarse ya sea que la roca sea basalto o granito. Vuelos sobre regiones de composición conocida han arrojado la siguiente información:

	Graníticas	**Basálticas**
$R_1 < R_2 < R_3$	60%	10%
$R_1 < R_3 < R_2$	25%	20%
$R_3 < R_1 < R_2$	15%	70%

Suponga que para una roca seleccionada al azar en cierta región, $P(\text{granito}) = .25$ y $P(\text{basalto}) = .75$.

a. Demuestre que $P(\text{granito} \mid R_1 < R_2 < R_3) > P(\text{basalto} \mid R_1 < R_2 < R_3)$. Si las mediciones dieron $R_1 < R_2 < R_3$, ¿clasificaría la roca como granito o como basalto?

b. Si las mediciones dieron $R_1 < R_3 < R_2$, ¿cómo clasificaría la roca? Responda la misma pregunta para $R_3 < R_1 < R_2$.

c. Con las reglas de clasificación indicadas en los incisos (a) y (b) cuando se selecciona una roca de esta región, ¿cuál es la probabilidad de una clasificación errónea? [*Sugerencia*: G podría ser clasificada como B o B como G y $P(B)$ y $P(G)$ son conocidas.]

d. Si $P(\text{granito}) = p$ en lugar de .25, ¿existen valores de p (aparte de 1) para los cuales una roca siempre sería clasificada como granito?

107. A un sujeto se le permite una secuencia de vistazos para detectar un objetivo. Sea $G_i = \{\text{el objetivo es detectado en el vistazo } i\text{-ésimo}\}$, con $p_i = P(G_i)$. Suponga que los G_i son eventos independientes y escriba una expresión para la probabilidad de que el objetivo haya sido detectado al final del vistazo n-ésimo. [*Nota*: este modelo se discute en "Predicting Aircraft Detectability", *Human Factors*, 1979: 277–291.]

108. En un juego de beisbol de ligas pequeñas, el lanzador del equipo A lanza un "strike" 50% del tiempo y una bola 50% del tiempo; los lanzamientos sucesivos son independientes unos de otros y el lanzador nunca golpea a un bateador. Sabiendo esto, el mánager del equipo B ha instruido al primer bateador para que no le batee a nada.

Calcule la probabilidad de que

a. El bateador reciba base por bolas en el cuarto lanzamiento

b. El bateador reciba base por bolas en el sexto lanzamiento (por lo que dos de los primeros cinco deben ser "strikes"), por medio de un argumento de conteo o un diagrama de árbol.

c. El bateador recibe base por bolas.

d. El primer bateador anota mientras no hay ningún "out" (suponiendo que cada bateador utiliza la estrategia de no batearle a nada)

109. Cuatro ingenieros, A, B, C y D han sido citados para entrevistas de trabajo a las 10 a.m. el viernes 13 de enero, en Random Sampling, Inc. El gerente de personal ha programado a los cuatro para las oficinas de entrevistas 1, 2, 3 y 4, respectivamente. Sin embargo, el secretario del gerente no está enterado de esto, por lo que los asigna a las oficinas de un modo completamente aleatorio (¡qué más!) ¿Cuál es la probabilidad de que

a. los cuatro terminen en la oficina correcta?

b. ninguno de los cuatro termine en la oficina correcta?

110. Una aerolínea particular opera vuelos a las 10 a.m. de Chicago a Nueva York, Atlanta y Los Ángeles. Sea A el evento en que el vuelo a Nueva York está lleno y defina los eventos B y C en forma análoga para los otros dos vuelos. Suponga que $P(A) = .6$, $P(B) = .5$, $P(C) = .4$ y los tres eventos son independientes. ¿Cuál es la probabilidad de que

a. los tres vuelos estén llenos? ¿Que por lo menos uno no esté lleno?

b. sólo el vuelo a Nueva York esté lleno? ¿Que exactamente uno de los tres vuelos esté lleno?

111. Un gerente de personal va a entrevistar a cuatro candidatos para un puesto. Éstos están clasificados como 1, 2, 3 y 4 en orden de preferencia, y serán entrevistados en orden aleatorio. Sin embargo, al final de cada entrevista, el gerente sabrá sólo cómo se compara el candidato actual con los candidatos previamente entrevistados. Por ejemplo, el orden de entrevista 3, 4, 1, 2 no genera información después de la primera entrevista, muestra que el segundo candidato es peor que el primero y que el tercero es mejor que los primeros dos. Sin embargo, el orden 3, 4, 2, 1 generaría la misma información después de cada una de las primeras tres entrevistas. El gerente desea contratar al mejor candidato pero debe tomar una decisión irrevocable de contratarlo o no contratarlo después de cada entrevista. Considere la siguiente estrategia: rechazar automáticamente al primer candidato s y luego contratar al primer candidato subsiguiente que resulte mejor entre los que ya fueron entrevistados (si tal candidato no aparece, el último entrevistado es el contratado).

Por ejemplo, con $s = 2$, el orden 3, 4, 1, 2 permitiría contratar al mejor, en tanto que el orden 3, 1, 2, 4, no. De los cuatro valores s posibles (0, 1, 2 y 3), ¿cuál incrementa al máximo a P(el mejor es contratado)? [*Sugerencia*: escriba los 24 ordenamientos de entrevista igualmente probables: $s = 0$ significa que el primer candidato es automáticamente contratado.]

112. Considere cuatro eventos independientes A_1, A_2, A_3 y A_4, y sea $p_i = P(A_i)$ con $i = 1, 2, 3, 4$. Exprese la probabilidad de que por lo menos uno de estos eventos ocurra en función de las p_i y haga lo mismo para la probabilidad de que por lo menos dos de los eventos ocurran.

113. Una caja contiene los siguientes cuatro papelitos y cada uno tiene exactamente las mismas dimensiones: (1) gana el premio 1: (2) gana el premio 2: (3) gana el premio 3; (4) gana los premios 1, 2, y 3. Se selecciona un papelito al azar. Sea $A_1 = \{$gana el premio 1$\}$, $A_2 = \{$gana el premio 2$\}$ y $A_3 = \{$gana el premio 3$\}$. Demuestre que A_1 y A_2 son independientes, que A_1 y A_3 son independientes, y que A_2 y A_3 también son independientes (ésta es una independencia *por pares*). Sin embargo, demuestre que $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$, así que los tres eventos *no* son mutuamente independientes.

114. Demuestre que si A_1, A_2 y A_3 son eventos independientes, entonces $P(A_1 \mid A_2 \cap A_3) = P(A_1)$.

Bibliografía

Durrett, Richard, *Elementary Probability for Applications*, Cambridge Univ. Press, Londres, Inglaterra, 2009. Una presentación concisa a un nivel un poco más alto que este texto.

Mosteller, Frederick, Robert Rourke y George Thomas, *Probability with Statistical Applications* (2a. ed.), Addison-Wesley, Reading, MA, 1970. Una muy buena introducción a la probabilidad, con muchos ejemplos entretenidos; especialmente buenos con respecto a reglas de conteo y su aplicación.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Application* (2a. ed.), Macmillan, Nueva York, 1994. Una amplia introducción a la probabilidad escrita a un nivel matemático un poco más alto que este texto pero que contiene muchos buenos ejemplos.

Ross, Sheldon, *A First Course in Probability* (6a. ed.), Macmillan, Nueva York, 2009. Algo concisamente escrito y más matemáticamente complejo que este texto pero contiene una gran cantidad de ejemplos y ejercicios interesantes.

Winkler, Robert, *Introduction to Bayesian Inference and Decision*, Holt, Rinehart & Winston, Nueva York, 1972. Una muy buena introducción a la probabilidad subjetiva.

3 Variables aleatorias discretas y distribuciones de probabilidad

INTRODUCCIÓN

Ya sea que un experimento produzca resultados cualitativos o cuantitativos, los métodos de análisis estadístico requieren enfocarse en ciertos aspectos numéricos de los datos (como la proporción muestral x/n, la media $\bar{x}$ o la desviación estándar σ). El concepto de variable aleatoria permite pasar de los resultados experimentales a la función numérica de los resultados. Existen dos tipos fundamentalmente diferentes de variables aleatorias: las variables aleatorias discretas y las variables aleatorias continuas. En este capítulo se examinan las propiedades básicas y se discuten los ejemplos más importantes de variables discretas. El capítulo 4 se enfoca en las variables aleatorias continuas.

3.1 Variables aleatorias

En cualquier experimento existen numerosas características que pueden ser observadas o medidas, pero en la mayoría de los casos un experimentador se enfoca en algún aspecto específico o aspectos de una muestra. Por ejemplo, en un estudio de patrones de viaje entre los suburbios y la ciudad en un área metropolitana, a cada individuo en una muestra se le podría preguntar sobre la distancia que recorre para ir de su casa al trabajo y viceversa, y el número de personas que viajan en el mismo vehículo, pero no sobre su coeficiente intelectual, ingreso, tamaño de su familia y otras características. Por otra parte, un investigador puede probar una muestra de componentes y anotar sólo el número de los que han fallado dentro de 1000 horas, en lugar de anotar los tiempos de falla individuales.

En general, cada resultado de un experimento puede ser asociado con un número especificando una regla de asociación (p. ej., el número entre la muestra de diez componentes que no duran 1000 horas o el peso total del equipaje en una muestra de 25 pasajeros de aerolínea). Semejante regla de asociación se llama **variable aleatoria**, variable porque diferentes valores numéricos son posibles y aleatoria porque el valor observado depende de cuál de los posibles resultados experimentales resulte (figura 3.1).

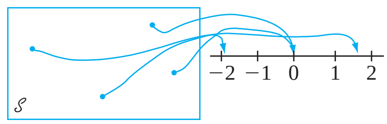


Figura 3.1 Una variable aleatoria

DEFINICIÓN

> Para un espacio muestral dado $\mathcal{S}$ de algún experimento, una **variable aleatoria** es cualquier regla que asocia un número con cada resultado en $\mathcal{S}$. En lenguaje matemático, una variable aleatoria es una función cuyo dominio es el espacio muestral y cuyo rango es el conjunto de los números reales.

Se acostumbra denotar las variables aleatorias con letras mayúsculas, tales como X y Y, que son las de cerca del final del alfabeto. En contraste al uso previo de una letra minúscula, tal como x, para denotar una variable, ahora se utilizarán letras minúsculas para representar algún valor particular de la variable aleatoria correspondiente. La notación $X(s) = x$ significa que x es el valor asociado con el resultado s por medio de la variable aleatoria X.

Ejemplo 3.1 Cuando un estudiante llama a un servicio de asistencia universitaria para apoyo técnico, él/ella podrá inmediatamente hablar con alguien (S) o será puesto en espera (F). Con $\mathcal{S} = \{S, F\}$, la variable aleatoria X se define como

$$X(S) = 1 \quad X(F) = 0$$

La variable aleatoria X indica si (1) o no (0) el estudiante puede hablar inmediatamente con alguien. ■

La variable aleatoria X en el ejemplo 3.1 se especificó al poner en lista explícitamente cada elemento de $\mathcal{S}$ y el número asociado. Una lista como ésa es tediosa si $\mathcal{S}$ contiene más de algunos cuantos resultados, pero con frecuencia puede ser evitada.

Ejemplo 3.2 Considere el experimento en el cual se marca un número telefónico en cierto código de área con un marcador de números aleatorio (tales dispositivos los utilizan en forma extensa en organizaciones encuestadoras) y defina una variable aleatoria Y como

$$Y = \begin{cases} 1 & \text{si el número seleccionado no aparece en el directorio} \\ 0 & \text{si el número seleccionado sí aparece en el directorio} \end{cases}$$

Por ejemplo, si 5282966 aparece en el directorio telefónico, entonces $Y(5282966) = 0$ en tanto que $Y(7727350) = 1$ dice que el número 7727350 no aparece en el directorio telefónico. Una descripción en palabras de esta índole es más económica que una lista completa, por lo que se utilizará tal descripción siempre que sea posible. ■

En los ejemplos 3.1 y 3.2, los únicos valores posibles de la variable aleatoria fueron 0 y 1. Tal variable aleatoria se presenta con suficiente frecuencia como para darle un nombre especial, en honor del individuo que la estudió primero.

DEFINICIÓN

Cualquier variable aleatoria cuyos únicos valores posibles son 0 y 1 se llama **variable aleatoria de Bernoulli.**

En ocasiones se deseará definir y estudiar diferentes variables del mismo espacio muestral.

Ejemplo 3.3 El ejemplo 2.3 describe un experimento en el cual se determinó el número de bombas en uso en cada una de dos gasolineras. Defina las variables aleatorias X, Y y U como

$X =$ el número total de bombas en uso en las dos gasolineras

$Y =$ la diferencia entre el número de bombas en uso en la gasolinera 1 y el número en uso en la gasolinera 2

$U =$ el máximo de los números de bombas en uso en las dos gasolineras

Si se realiza este experimento y se obtiene $s = (2, 3)$, entonces $X((2, 3)) = 2 + 3 = 5$, por lo que se dice que el valor observado de X fue $x = 5$. Asimismo, el valor observado de Y sería $y = 2 - 3 = -1$ y el de U sería $u =$ máx $(2, 3) = 3$. ■

Cada una de las variables aleatorias de los ejemplos 3.1–3.3 puede asumir sólo un número finito de posibles valores. Éste no tiene que ser el caso.

Ejemplo 3.4 Se considera un experimento en que se examinaron baterías de 9 volts hasta que se obtuvo una con un voltaje aceptable (S). El espacio muestral es $\mathcal{S} = \{S, FS, FFS, \ldots\}$. Defina una variable aleatoria X como

$X =$ el número de baterías examinadas antes de que se termine el experimento

En ese caso $X(S) = 1$, $X(FS) = 2$, $X(FFS) = 3, \ldots, X(FFFFFFS) = 7$, y así sucesivamente. Cualquier entero positivo es un valor positivo de X, así que el conjunto de valores posibles es infinito. ■

Ejemplo 3.5 Suponga que del mismo modo aleatorio se selecciona un lugar (latitud y longitud) en el territorio continental de Estados Unidos. Defina una variable aleatoria Y como

$Y =$ la altura sobre el nivel del mar en el lugar seleccionado

Por ejemplo, si el lugar seleccionado fuera (39°50′N, 98°35′O), entonces se podría tener $Y((39°50'N, 98°35'O)) = 1748.26$ pies. El valor más grande posible de Y es 14,494 (Monte Whitney) y el valor más pequeño posible es -282 (Valle de la Muerte). El conjunto de todos los valores posibles de Y es el conjunto de todos los números en el intervalo entre -282 y 14,494, es decir,

$$\{y: y \text{ es un número}, -282 \le y \le 14{,}494\}$$

y existe un número infinito de números en este intervalo. ■

Dos tipos de variables aleatorias

En la sección 1.2, se distinguió entre los datos que resultan de observaciones de una variable de conteo y los datos obtenidos observando valores de una variable de medición. Una distinción un poco más formal caracteriza dos tipos diferentes de variables aleatorias.

DEFINICIÓN

Una variable aleatoria **discreta** es una variable aleatoria cuyos valores posibles constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita en la cual existe un primer elemento, un segundo elemento, y así sucesivamente ("contablemente" infinita).

Una variable aleatoria es **continua** si *ambas* de las siguientes condiciones se cumplen:

1. Su conjunto de valores posibles se compone de todos los números que hay en un solo intervalo sobre la línea de numeración (posiblemente de extensión infinita, es decir, desde $-\infty$ hasta ∞) o todos los números en una unión disjunta de dichos intervalos (por ejemplo, $[0, 10] \cup [20, 30]$).
2. Ningún valor posible de la variable tiene probabilidad positiva, esto es, $P(X = c) = 0$ con cualquier valor posible de c.

Aunque cualquier intervalo sobre la línea de numeración contiene un número infinito de números, se puede demostrar que no existe ninguna forma de crear una lista infinita de todos estos valores, pues existen demasiados de ellos. La segunda condición que describe una variable aleatoria continua es tal vez contraintuitiva, puesto que parecería que implica una probabilidad total de cero para todos los valores posibles. Pero en el capítulo 4 se verá que los *intervalos* de valores tienen probabilidad positiva; la probabilidad de un intervalo se reducirá a cero a medida que su ancho tienda a cero.

Ejemplo 3.6 Todas las variables aleatorias de los ejemplos 3.1–3.4 son discretas. Como otro ejemplo, suponga que se eligen al azar parejas de casados y que a cada persona se le hace una prueba de sangre hasta encontrar un esposo y esposa con el mismo factor Rh. Con X = el número de pruebas de sangre que serán realizadas, los posibles valores de X son $D = \{2, 4, 6, 8, \ldots\}$. Como los posibles valores se dieron en secuencia, X es una variable aleatoria discreta. ■

Para estudiar las propiedades básicas de las variables aleatorias discretas, sólo se requieren las herramientas de las matemáticas discretas: sumas y diferencias. El estudio de variables continuas requiere las matemáticas continuas del cálculo: integrales y derivadas.

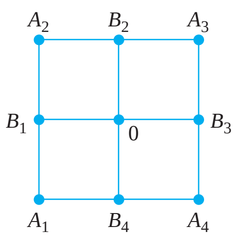
EJERCICIOS Sección 3.1 (1–10)

1. Una viga de concreto puede fallar por esfuerzo cortante (S) o por flexión (F). Suponga que se seleccionan al azar tres vigas que fallaron y se determina el tipo de falla de cada una. Sea X = el número de vigas entre las tres seleccionadas que fallaron por esfuerzo cortante. Ponga en lista cada resultado en el espacio muestral junto con el valor asociado de X.
2. Dé tres ejemplos de variables aleatorias de Bernoulli (aparte de los que aparecen en el texto).
3. Con el experimento del ejemplo 3.3, defina dos variables aleatorias más y mencione los valores posibles de cada una.
4. Sea X = el número de dígitos no cero en un código postal seleccionado al azar. ¿Cuáles son los posibles valores de X? Dé tres posibles resultados y sus valores X asociados.
5. Si el espacio muestral $\mathcal{S}$ es un conjunto infinito, ¿implica esto necesariamente que cualquier variable aleatoria X definida a partir de $\mathcal{S}$ tendrá un conjunto infinito de posibles valores? Si es sí, por qué. Si no, dé un ejemplo.
6. A partir de una hora fija, cada carro que entra a una intersección es observado para ver si da vuelta a la izquierda (I), a la derecha (D) o si sigue de frente (F). El experimento termina en cuanto se observa que un carro da vuelta a la izquierda. Sea X = el número de carros observados. ¿Cuáles son los posibles valores de X? Dé cinco resultados y sus valores X asociados.
7. Para cada variable aleatoria definida aquí, describa el conjunto de posibles valores de la variable y diga si la variable es discreta.

a. X = el número de huevos no quebrados en una caja de huevos estándar seleccionada al azar.
b. Y = el número de estudiantes en una lista de clase de un curso particular que no asisten el primer día de clases.
c. U = el número de veces que un aprendiz tiene que hacerle *swing* a una pelota de golf antes de golpearla.
d. X = la longitud de una serpiente de cascabel seleccionada en forma aleatoria.
e. Z = la cantidad de regalías devengada por la venta de la primera edición de 10,000 libros de texto.
f. Y = el pH de una muestra de suelo elegida al azar.
g. X = la tensión (lb/pulg^2) a la cual una raqueta de tenis seleccionada al azar fue encordada.
h. X = el número total de lanzamientos al aire de una moneda requerido para que tres individuos obtengan una coincidencia (*AAA* o *SSS*).

8. Cada vez que un componente se somete a prueba, ésta es un éxito (S) o una falla (F). Suponga que el componente se prueba repetidamente hasta que ocurre un éxito en tres pruebas *consecutivas*. Sea Y el número necesario de pruebas para lograrlo. Haga una lista de todos los resultados correspondientes a los cinco posibles valores más pequeños de Y y diga qué valor de Y está asociado con cada uno.

9. Un individuo de nombre Claudius se encuentra en el punto 0 del diagrama adjunto.



Con un dispositivo de aleatorización apropiado (tal como un dado tetraédrico, uno que tiene cuatro lados), Claudius primero se mueve a uno de los cuatro lugares B_1, B_2, B_3, B_4. Una vez que está en uno de estos lugares, se utiliza otro dispositivo de aleatorización para decidir si Claudius regresa a 0 o visita uno de los otros dos lugares adyacentes. Este proceso continúa entonces; después de cada movimiento, se determina otro movimiento a uno de los (nuevos) puntos adyacentes lanzando al aire un dado o moneda apropiada.
a. Sea X = el número de movimientos que Claudius hace antes de regresar a 0. ¿Cuáles son los posibles valores de X? ¿Es X discreta o continua?
b. Si también se permiten movimientos a lo largo de los trayectos diagonales que conectan 0 con A_1, A_2, A_3 y A_4, respectivamente, responda las preguntas del inciso (a).

10. Se determinará el número de bombas en uso tanto en la gasolinera de seis bombas como en la gasolinera de cuatro bombas. Dé los posibles valores de cada una de las siguientes variables aleatorias:
a. T = el número total de bombas en uso
b. X = la diferencia entre el número en uso en las gasolineras 1 y 2
c. U = el número máximo de bombas en uso en una u otra gasolinera
d. Z = el número de gasolineras que tienen exactamente dos bombas en uso

3.2 Distribuciones de probabilidad para variables aleatorias discretas

Las probabilidades asignadas a varios resultados en $\mathcal{S}$ determinan a su vez las probabilidades asociadas con los valores de cualquier variable aleatoria X particular. La *distribución de probabilidad de X* dice cómo está distribuida (asignada) la probabilidad total de 1 entre los varios posibles valores de X. Supóngase, por ejemplo, que una empresa acaba de adquirir cuatro impresoras láser y sea X el número de éstas que requieren servicio durante el periodo de garantía. Los posibles valores de X son entonces 0, 1, 2, 3 y 4. La distribución de probabilidad dirá cómo está subdividida la probabilidad de 1 entre estos cinco posibles valores: cuánta probabilidad está asociada con el valor 0 de X, cuánta está adjudicada al valor 1 de X, y así sucesivamente. Se utilizará la siguiente notación para las probabilidades en la distribución:

$$p(0) = \text{la probabilidad del valor 0 de } X = P(X = 0)$$
$$p(1) = \text{la probabilidad del valor 1 de } X = P(X = 1)$$

y así sucesivamente. En general, $p(x)$ denotará la probabilidad asignada al valor de x.

Ejemplo 3.7 El Departamento de Estadística de Cal Poly tiene un laboratorio con seis computadoras reservadas para estudiantes de estadística. Sea X el número de computadoras que están en servicio a una hora particular del día. Suponga que la distribución de probabilidad de X es como se da en la tabla siguiente; la primera fila de la tabla contiene los posibles valores de X y la segunda da la probabilidad de dicho valor.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

Ahora se pueden usar propiedades de probabilidad elemental para calcular otras probabilidades de interés. Por ejemplo, la probabilidad de que cuando mucho 2 computadoras estén en servicio es

$$P(X \leq 2) = P(X = 0 \text{ o } 1 \text{ o } 2) = p(0) + p(1) + p(2) = .05 + .10 + .15 = .30$$

Como el evento de que *por lo menos 3 computadoras estén en servicio* es complementario a *cuando mucho dos computadoras están en servicio,*

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - .30 = .70$$

la que, desde luego, también se obtiene sumando las probabilidades de los valores 3, 4, 5 y 6. La probabilidad de que entre 2 y 5 computadoras *inclusive* estén en servicio es

$$P(2 \leq X \leq 5) = P(X = 2, 3, 4 \text{ o } 5) = .15 + .25 + .20 + .15 = .75$$

en tanto que la probabilidad de que el número de computadoras en servicio esté *estrictamente entre* 2 y 5 es

$$P(2 < X < 5) = P(X = 3 \text{ o } 4) = .25 + .20 = .45$$

■

DEFINICIÓN

La **distribución de probabilidad** o **función de masa de probabilidad** (fmp) de una variable discreta se define para cada número x como $p(x) = P(X = x) = P$ (todas las $s \in \mathcal{S}$: $X(s) = x$).

En palabras, para cada valor posible x de la variable aleatoria, la función de masa de probabilidad especifica la probabilidad de observar dicho valor cuando se realiza el experimento. Se requieren las condiciones $p(x) \geq 0$ y $\sum_{\text{todas las } x \text{ posibles}} p(x) = 1$ de cualquier función de masa de probabilidad.

La función de masa de probabilidad de X en el ejemplo previo se dio simplemente en la descripción del problema. A continuación se consideran varios ejemplos en los cuales se explotan varias propiedades de probabilidad para obtener la distribución deseada.

Ejemplo 3.8 Seis lotes de componentes están listos para ser enviados por un proveedor. El número de componentes defectuosos en cada lote es como sigue:

Lote	1	2	3	4	5	6
Número de componentes defectuosos	0	2	0	1	2	0

Uno de estos lotes tiene que ser seleccionado al azar para ser enviado a un cliente particular. Sea X el número de defectuosos en el lote seleccionado. Los tres posibles valores de X son 0, 1 y 2. De los seis eventos simples igualmente probables, tres dan por resultado $X = 0$, uno $X = 1$ y los otros dos $X = 2$. Entonces

$$p(0) = P(X = 0) = P(\text{el lote 1 o 3 o 6 es enviado}) = \frac{3}{6} = .500$$

$$p(1) = P(X = 1) = P(\text{el lote 4 es enviado}) = \frac{1}{6} = .167$$

$$p(2) = P(X = 2) = P(\text{el lote 2 o 5 es enviado}) = \frac{2}{6} = .333$$

Es decir, una probabilidad de .500 se asigna al valor 0 de X, una probabilidad de .167 se asigna al valor 1 de X, y la probabilidad restante, .333, se asocia con el valor 2 de X. Los valores de X junto con sus probabilidades especifican la función de masa de probabilidad. Si este experimento se repitiera una y otra vez, a la larga $X = 0$ ocurriría la mitad del tiempo, $X = 1$ un sexto del tiempo y $X = 2$ un tercio del tiempo. ■

Ejemplo 3.9 Considere si la siguiente persona que compre una computadora en cierta tienda de electrónicos elegirá un modelo portátil o uno de escritorio. Sea

$$X = \begin{cases} 1 & \text{si el cliente compra una computadora de escritorio} \\ 0 & \text{si el cliente compra una computadora portátil} \end{cases}$$

Si 20% de todos los compradores durante esa semana seleccionan una de escritorio, la función de masa de probabilidad de X es

$$p(0) = P(X = 0) = P(\text{el siguiente cliente compra un modelo portátil}) = .8$$
$$p(1) = P(X = 1) = P(\text{el siguiente cliente compra un modelo de escritorio}) = .2$$
$$p(x) = P(X = x) = 0 \text{ con } x \neq 0 \text{ o } 1$$

Una descripción equivalente es

$$p(x) = \begin{cases} .8 & \text{si } x = 0 \\ .2 & \text{si } x = 1 \\ 0 & \text{si } x \neq 0 \text{ o } 1 \end{cases}$$

La figura 3.2 es una ilustración de esta función de masa de probabilidad, llamada *gráfica lineal*. X es, desde luego, una variable aleatoria de Bernoulli y $p(x)$ es una función de masa de probabilidad de Bernoulli.

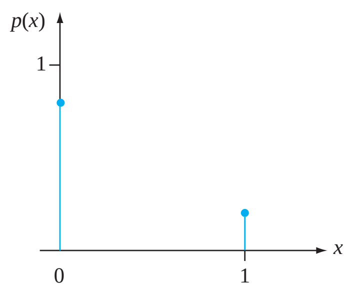


Figura 3.2 Gráfica lineal para la función de masa de probabilidad de Bernoulli en el ejemplo 3.9 ■

Ejemplo 3.10 Considere un grupo de cinco donadores de sangre potenciales, a, b, c, d y e, de los cuales sólo a y b tienen sangre tipo O+. Se determinará en orden aleatorio el tipo de sangre con cinco muestras, una de cada individuo, hasta que se identifique un individuo O+. Sea la variable aleatoria Y = el número de exámenes de sangre para identificar un individuo O+. Entonces la función de masa de probabilidad de Y es

$$p(1) = P(Y = 1) = P(a \text{ o } b \text{ examinados primero}) = \frac{2}{5} = .4$$

$$p(2) = P(Y = 2) = P(c, d \text{ o } e \text{ primero, y luego } a \text{ o } b)$$

$$= P(c, d, \text{ o } e \text{ primero}) \cdot P(a \text{ o } b \text{ a continuación} \mid c, d \text{ o } e \text{ primero}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$

$$p(3) = P(Y = 3) = P(c, d \text{ o } e \text{ primero y segundo, y luego } a \text{ o } b)$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = .2$$

$$p(4) = P(Y = 4) = P(c, d \text{ y } e \text{ primero}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = .1$$

$$p(y) = 0 \quad \text{si } y \neq 1, 2, 3, 4$$

En forma tabular, la función de masa de probabilidad es

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

donde cualquier valor de y que no aparece en la tabla recibe cero probabilidad. La figura 3.3 muestra una gráfica lineal de la función de masa de probabilidad.

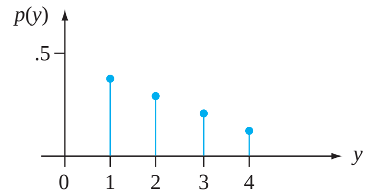


Figura 3.3 Gráfica lineal para la función de masa de probabilidad de Bernoulli del ejemplo 3.10 ■

Un modelo utilizado en física para un sistema de "masas puntuales" sugirió el nombre "función de masa de probabilidad". En este modelo, las masas están distribuidas en varios lugares x a lo largo de un eje unidimensional. La función de masa de probabilidad describe cómo está distribuida la masa de probabilidad total de 1 en varios puntos a lo largo del eje de posibles valores de la variable aleatoria (dónde y cuánta masa hay en cada x).

Otra representación pictórica útil de una función de masa de probabilidad, llamada **histograma de probabilidad,** es similar a los histogramas discutidos en el capítulo 1. Sobre cada y con $p(y) > 0$, se construye un rectángulo con su centro en y. La altura de cada rectángulo es proporcional a $p(y)$ y la base es la misma para todos los rectángulos. Cuando los valores posibles están equidistantes, con frecuencia se selecciona la base como la distancia entre valores y sucesivos (aunque podría ser más pequeña). La figura 3.4 muestra dos histogramas de probabilidad.

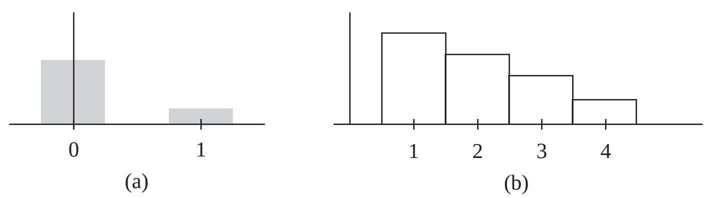


Figura 3.4 Histogramas de probabilidad: (a) Ejemplo 3.9; (b) Ejemplo 3.10

A menudo es útil pensar en una función de masa de probabilidad como un modelo matemático de una población discreta.

Ejemplo 3.11 Considere seleccionar al azar un estudiante de entre los 15,000 inscritos en el semestre actual en la Universidad Mega. Sea X = el número de cursos en los cuales el estudiante seleccionado está inscrito y suponga que X tiene la siguiente función de masa de probabilidad:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

Una forma de ver esta situación es pensar en la población como compuesta de 15,000 individuos, cada uno con su propio valor X; la proporción con cada valor de X está dada por $p(x)$. Un punto de vista alternativo es olvidarse de los estudiantes y pensar en la población como compuesta de los valores X: existen algunos 1 en la población, algunos 2, . . . , y finalmente algunos 7. La población se compone entonces de los números 1, 2, . . . , 7 (por lo tanto es discreta) y $p(x)$ da un modelo para la distribución de los valores de población. ■

Una vez que se tiene el modelo de la población, se utilizará para calcular valores de características de la población (p. ej., la media μ) y para hacer inferencias sobre tales características.

Parámetro de una distribución de probabilidad

La función de masa de probabilidad de Bernoulli en el ejemplo 3.9, fue $p(0) = .8$ y $p(1) = .2$ porque 20% de todos los compradores seleccionaron una computadora de escritorio. En otro almacén, puede ser el caso que $p(0) = .9$ y $p(1) = .1$. Más generalmente, la función de masa de probabilidad de cualquier variable aleatoria de Bernoulli puede ser expresada en la forma $p(1) = \alpha$ y $p(0) = 1 - \alpha$, donde $0 < \alpha < 1$. Como la función de masa de probabilidad depende del valor particular de α, con frecuencia se escribe $p(x; \alpha)$ en lugar de sólo $p(x)$:

$$p(x; \alpha) = \begin{cases} 1 - \alpha & \text{si } x = 0 \\ \alpha & \text{si } x = 1 \\ 0 & \text{de lo contrario} \end{cases} \tag{3.1}$$

Entonces cada opción de α en la expresión (3.1) da una función de masa de probabilidad diferente.

DEFINICIÓN

Supóngase que $p(x)$ depende de la cantidad que puede ser asignada a cualquiera de un número de valores posibles, y cada valor determina una distribución de probabilidad diferente. Tal cantidad se llama **parámetro** de distribución. El conjunto de todas las distribuciones de probabilidad para diferentes valores del parámetro se llama **familia** de distribuciones de probabilidad.

La cantidad α en la expresión (3.1) es un parámetro. Cada número diferente α entre 0 y 1 determina un miembro diferente de la familia de distribuciones de Bernoulli.

Ejemplo 3.12 A partir de un tiempo fijo, se observa el sexo de cada niño recién nacido en un hospital hasta que nace un varón (B). Sea $p = P(B)$ y suponga que los nacimientos sucesivos son independientes y defina la variable aleatoria X como $x =$ número de nacimientos observados. Entonces

$$p(1) = P(X = 1) = P(B) = p$$
$$p(2) = P(X = 2) = P(GB) = P(G) \cdot P(B) = (1 - p)p$$

y

$$p(3) = P(X = 3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1 - p)^2 p$$

Continuando de esta manera, emerge una fórmula general:

$$p(x) = \begin{cases} (1 - p)^{x-1} p & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases} \tag{3.2}$$

El parámetro p puede asumir cualquier valor entre 0 y 1. La expresión (3.2) describe la familia de distribuciones *geométricas*. En el ejemplo del sexo, $p = .51$ podría ser apropiado, pero si estábamos buscando el primer hijo con sangre Rh positiva, entonces podríamos tener $p = .85$. ■

Función de distribución acumulativa

Para algún valor fijo x, a menudo se desea calcular la probabilidad de que el valor observado de X será cuando mucho x. Por ejemplo, la función de masa de probabilidad en el ejemplo 3.8 fue

$$p(x) = \begin{cases} .500 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & \text{de lo contrario} \end{cases}$$

La probabilidad de que X sea cuando mucho 1 es entonces

$$P(X \leq 1) = p(0) + p(1) = .500 + .167 = .667$$

En este ejemplo, $X \leq 1.5$ si y sólo si $X \leq 1$, por lo tanto

$$P(X \leq 1.5) = P(X \leq 1) = .667$$

Asimismo,

$$P(X \leq 0) = P(X = 0) = .5, \quad P(X \leq .75) = .5$$

Y de hecho con cualquier x que satisfaga $0 \leq x < 1$, $P(X \leq x) = .5$. El valor de X más grande posible es 2, por lo tanto

$$P(X \leq 2) = 1, \quad P(X \leq 3.7) = 1, \quad P(X \leq 20.5) = 1$$

y así sucesivamente. Obsérvese que $P(X < 1) < P(X \leq 1)$ puesto que la segunda parte de la desigualdad incluye la probabilidad del valor 1 de X, en tanto que la primera no. Más generalmente, cuando X es discreta y x es un valor posible de la variable, $P(X < x) < P(X \leq x)$.

DEFINICIÓN

La **función de distribución acumulativa** (fda) $F(x)$ de una variable aleatoria discreta X con función de masa de probabilidad $p(x)$ se define para cada número x como

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y) \tag{3.3}$$

Para cualquier número x, $F(x)$ es la probabilidad de que el valor observado de X será cuando mucho x.

Ejemplo 3.13 Una tienda vende unidades de memoria flash, ya sea con 1 GB, 2 GB, 4 GB, 8 GB o 16 GB de memoria. La tabla adjunta muestra la distribución de Y = la cantidad de memoria en un disco comprado:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

Primero se determina $F(y)$ para cada uno de los cinco valores posibles de Y:

$$\begin{aligned} F(1) &= P(Y \leq 1) = P(Y = 1) = p(1) = .05 \\ F(2) &= P(Y \leq 2) = P(Y = 1 \text{ o } 2) = p(1) + p(2) = .15 \\ F(4) &= P(Y \leq 4) = P(Y = 1 \text{ o } 2 \text{ o } 4) = p(1) + p(2) + p(4) = .50 \\ F(8) &= P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = .90 \\ F(16) &= P(Y \leq 16) = 1 \end{aligned}$$

Ahora con cualquier otro número y, $F(y)$ será igual al valor de F en el valor más próximo posible de Y a la izquierda de y. Por ejemplo,

$$F(2.7) = P(Y \leq 2.7) = P(Y \leq 2) = F(2) = .15$$
$$F(7.999) = P(Y \leq 7.999) = P(Y \leq 4) = F(4) = .50$$

Si y es menor que 1, $F(y) = 0$ [por ejemplo, $F(.58) = 0$], y si y es por lo menos 16, $F(y) = 1$ [por ejemplo, $F(25) = 1$]. La fda es, pues,

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \le y < 2 \\ .15 & 2 \le y < 4 \\ .50 & 4 \le y < 8 \\ .90 & 8 \le y < 16 \\ 1 & 16 \le y \end{cases}$$

En la figura 3.5 se muestra una gráfica de esta fda.

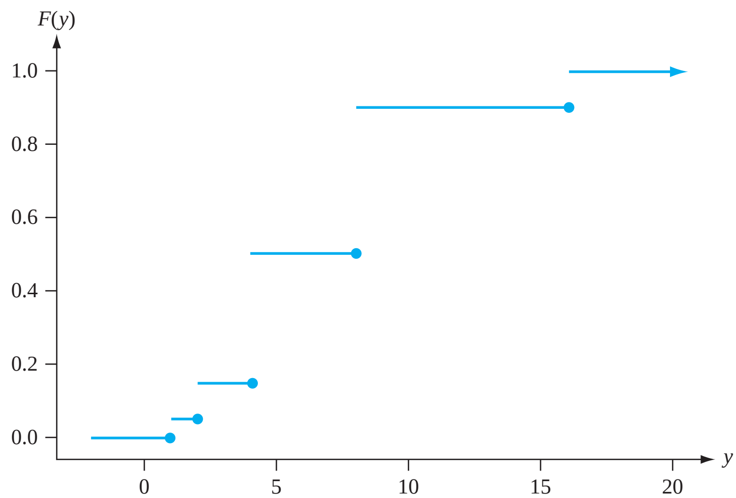


Figura 3.5 Gráfica de la función de distribución acumulativa del ejemplo 3.13 ■

Para una variable aleatoria discreta X, la gráfica de $F(x)$ mostrará un salto con cada valor posible de X, y será plana entre los valores posibles. Tal gráfica se conoce como **función escalón.**

Ejemplo 3.14 (Continuación del ejemplo 3.12)

La fmp de X = *el número de nacimientos* tenía la forma

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

Para cualquier entero positivo x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1}p = p\sum_{y=0}^{x-1}(1-p)^y \tag{3.4}$$

Para evaluar esta suma, se utiliza el hecho de que la suma parcial de una serie geométrica es

$$\sum_{y=0}^{k} a^y = \frac{1-a^{k+1}}{1-a}$$

Utilizando esto en la ecuación (3.4), con $a = 1 - p$ y $k = x - 1$, se obtiene

$$F(x) = p \cdot \frac{1 - (1 - p)^x}{1 - (1 - p)} = 1 - (1 - p)^x \quad x \text{ es un entero positivo}$$

Como F es una constante entre enteros positivos,

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - (1 - p)^{[x]} & x \geq 1 \end{cases} \tag{3.5}$$

donde $[x]$ es el entero más grande $\leq x$ (p. ej., $[2.7] = 2$). Así pues, si $p = .51$ como en el ejemplo de los nacimientos, entonces la probabilidad de tener que examinar cuando mucho cinco nacimientos para ver el primer varón es $F(5) = 1 - (.49)^5 = 1 - .0282 = .9718$, mientras que $F(10) \approx 1.0000$. Esta función de distribución acumulativa se ilustra en la figura 3.6.

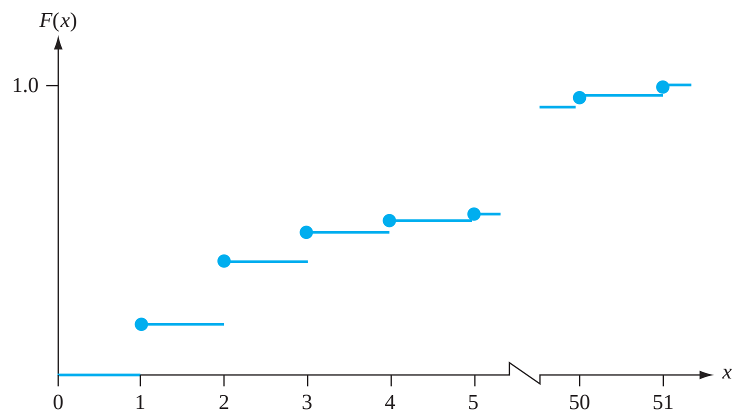


Figura 3.6 Gráfica de $F(x)$ para el ejemplo 3.14 ■

En los ejemplos presentados hasta ahora, la función de distribución acumulativa se dedujo de la función de masa de probabilidad. Este proceso puede ser invertido para obtener la función de masa de probabilidad a partir de la función de distribución acumulativa siempre que ésta esté disponible. Por ejemplo, considérese otra vez la variable aleatoria del ejemplo 3.7 (el número de computadoras usadas en un laboratorio); los valores posibles de X son $0, 1, \ldots, 6$. Entonces

$$\begin{aligned} p(3) &= P(X = 3) \\ &= [p(0) + p(1) + p(2) + p(3)] - [p(0) + p(1) + p(2)] \\ &= P(X \leq 3) - P(X \leq 2) \\ &= F(3) - F(2) \end{aligned}$$

Más generalmente, la probabilidad de que X quede dentro de un intervalo especificado es fácil de obtener a partir de la función de distribución acumulativa. Por ejemplo,

$$\begin{aligned} P(2 \leq X \leq 4) &= p(2) + p(3) + p(4) \\ &= [p(0) + \cdots + p(4)] - [p(0) + p(1)] \\ &= P(X \leq 4) - P(X \leq 1) \\ &= F(4) - F(1) \end{aligned}$$

Obsérvese que $P(2 \leq X \leq 4) \neq F(4) - F(2)$. Esto es porque el valor 2 de X está incluido en $2 \leq X \leq 4$, así que no se desea restar su probabilidad. Sin embargo, $P(2 < X \leq 4) = F(4) - F(2)$ porque $X = 2$ no está incluido en el intervalo $2 < X \leq 4$.

PROPOSICIÓN

Para dos números cualesquiera a y b con $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

donde "$a-$" representa el valor posible de X más grande que es estrictamente menor que a. En particular, si los únicos valores posibles son enteros, y si a y b son enteros, entonces

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a \text{ o } a + 1 \text{ o} \ldots \text{ o } b) \\ &= F(b) - F(a - 1) \end{aligned}$$

Con $a = b$ se obtiene $P(X = a) = F(a) - F(a - 1)$ en este caso.

La razón de restar $F(a-)$ en lugar de $F(a)$ es que se desea incluir $P(X = a)$; $F(b) - F(a)$ da $P(a < X \leq b)$. Esta proposición se utilizará extensamente cuando se calculen las probabilidades binomial y de Poisson en las secciones 3.4 y 3.6.

Ejemplo 3.15 Sea X = el número de días de ausencia por enfermedad tomados por un empleado seleccionado al azar de una gran compañía durante un año particular. Si el número máximo de días de ausencia por enfermedad permisibles al año es de 14, los valores posibles de X son 0, 1, . . . , 14. Con $F(0) = .58$, $F(1) = .72$, $F(2) = .76$, $F(3) = .81$, $F(4) = .88$ y $F(5) = .94$,

$$P(2 \leq X \leq 5) = P(X = 2, 3, 4 \text{ o } 5) = F(5) - F(1) = .22$$

y

$$P(X = 3) = F(3) - F(2) = .05$$

■

EJERCICIOS Sección 3.2 (11–28)

11. En un taller de servicio automotriz especializado en afinaciones se sabe que 45% de todas las afinaciones se realizan en automóviles de cuatro cilindros, 40% en automóviles de seis cilindros y 15% en automóviles de ocho cilindros. Sea X = el número de cilindros en el siguiente carro que va a ser afinado.
a. ¿Cuál es la función de masa de probabilidad de X?
b. Trace una gráfica lineal y un histograma de probabilidad de la función de masa de probabilidad del inciso (a).
c. ¿Cuál es la probabilidad de que el siguiente carro afinado sea de por lo menos seis cilindros? ¿Más de seis cilindros?

12. Las líneas aéreas en ocasiones venden boletos de más. Suponga que para un avión de 50 asientos, 55 pasajeros tienen boleto. Defina la variable aleatoria Y como el número de pasajeros con boleto que en realidad se presentan para el vuelo. La función de masa de probabilidad de Y aparece en la tabla adjunta.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	.05	.10	.12	.14	.25	.17	.06	.05	.03	.02	.01

a. ¿Cuál es la probabilidad de que el vuelo acomode a todos los pasajeros con boleto que se presenten?
b. ¿Cuál es la probabilidad de que no todos los pasajeros con boleto que aparecieron puedan ser acomodados?
c. Si usted es la primera persona en la lista de espera (lo que significa que será el primero en abordar el avión si hay boletos disponibles después de que todos los pasajeros con boleto hayan sido acomodados), ¿cuál es la probabilidad de que pueda tomar el vuelo? ¿Cuál es esta probabilidad si usted es la tercera persona en la lista de espera?

13. Una empresa de ventas en línea dispone de seis líneas telefónicas. Sea X el número de líneas en uso en un tiempo especificado. Suponga que la función de masa de probabilidad de X es la que se da en la tabla adjunta.

x	0	1	2	3	4	5	6
$p(x)$	.10	.15	.20	.25	.20	.06	.04

Calcule la probabilidad de cada uno de los siguientes eventos.
a. {cuando mucho tres líneas están en uso}
b. {menos de tres líneas están en uso}
c. {por lo menos tres líneas están en uso}
d. {entre dos y cinco líneas, inclusive, están en uso}
e. {entre dos y cuatro líneas, inclusive, no están en uso}
f. {por lo menos cuatro líneas no están en uso}

14. El departamento de planeación de un condado requiere que un contratista presente uno, dos, tres, cuatro o cinco formas (según la naturaleza del proyecto) para solicitar un permiso de construcción. Sea Y = número de formas requeridas del siguiente solicitante. Se sabe que la probabilidad de que se requieran y formas es proporcional a y, es decir, $p(y) = ky$ con $y = 1, \ldots, 5$.

a. ¿Cuál es el valor de k? [*Sugerencia*: $\sum_{y=1}^{5} p(y) = 1$.]

b. ¿Cuál es la probabilidad de que cuando mucho se requieran tres formas?

c. ¿Cuál es la probabilidad de que se requieran entre dos y cuatro formas (inclusive)?

d. ¿Podría ser $p(y) = y^2/50$ con $y = 1, \ldots, 5$ la función de masa de probabilidad de Y?

15. Muchos fabricantes cuentan con programas de control de calidad que incluyen la inspección de los materiales recibidos en busca de defectos. Suponga que un fabricante de computadoras recibe tarjetas madre en lotes de cinco. Se seleccionan dos tarjetas de cada lote para inspeccionarlas. Se puede representar los posibles resultados del proceso de selección por pares. Por ejemplo, el par (1, 2) representa la selección de las tarjetas 1 y 2 para inspección.

a. Mencione los diez posibles resultados diferentes.

b. Suponga que las tarjetas 1 y 2 son las únicas defectuosas en un lote de cinco. Dos tarjetas tienen que ser seleccionadas al azar. Defina X como el número de tarjetas defectuosas observadas entre las inspeccionadas. Encuentre la distribución de probabilidad de X.

c. Sea $F(x)$ la función de distribución acumulativa de X. Primero determine $F(0) = P(X \leq 0)$, $F(1)$ y $F(2)$; luego obtenga $F(x)$ para todas las demás x.

16. Algunas partes de California son particularmente propensas a los temblores. Suponga que en un área metropolitana, 25% de todos los propietarios de casa están asegurados contra daños provocados por terremotos. Se seleccionan al azar cuatro propietarios de casa; sea X el número entre los cuatro que están asegurados contra terremotos.

a. Encuentre la distribución de probabilidad de X. [*Sugerencia*: Sea S un propietario de casa asegurado y F uno no asegurado. Entonces un posible resultado es *SFSS*, con probabilidad (.25)(.75)(.25)(.25) y el valor 3 de X asociado. Existen otros 15 resultados.]

b. Trace el histograma de probabilidad correspondiente.

c. ¿Cuál es el valor más probable de X?

d. ¿Cuál es la probabilidad de que por lo menos dos de los cuatro seleccionados estén asegurados contra terremotos?

17. El voltaje de una batería nueva puede ser aceptable (A) o inaceptable (I). Una linterna requiere dos baterías, así que las baterías serán independientemente seleccionadas y probadas hasta encontrar dos aceptables. Suponga que 90% de todas las baterías tienen voltajes aceptables. Sea Y el número de baterías que deben ser probadas.

a. ¿Cuál es $p(2)$, es decir, $P(Y = 2)$?

b. ¿Cuál es $p(3)$? [*Sugerencia*: existen dos resultados diferentes que producen $Y = 3$.]

c. Para tener $Y = 5$, ¿qué debe ser cierto de la quinta batería seleccionada? Mencione los cuatro resultados con los cuales $Y = 5$ y luego determine $p(5)$.

d. Use el patrón de sus respuestas en los incisos (a)–(c) para obtener una fórmula general para $p(y)$.

18. Dos dados de seis caras son lanzados al aire en forma independiente. Sea M = el máximo de los dos lanzamientos (por lo tanto $M(1, 5) = 5$, $M(3, 3) = 3$, etcétera).

a. ¿Cuál es la función de masa de probabilidad de M? [*Sugerencia*: primero determine $p(1)$, luego $p(2)$, y así sucesivamente.]

b. Determine la función de distribución acumulativa de M y grafíquela.

19. Una biblioteca se suscribe a dos revistas diferentes de noticias semanales, cada una de las cuales se supone que llega en el correo de los miércoles. En realidad, cada una puede llegar el miércoles, jueves, viernes o sábado. Suponga que las dos llegan independientemente una de otra y para cada una $P(\text{mié}) = .3$, $P(\text{jue}) = .4$, $P(\text{vie}) = .2$ y $P(\text{sáb}) = .1$. Sea Y = el número de días después del miércoles que pasan para que ambas revistas lleguen (por lo tanto los posibles valores de Y son 0, 1, 2, o 3). Calcule la función de masa de probabilidad de Y. [*Sugerencia*: hay 16 posibles resultados: $Y(M, M) = 0$, $Y(V, J) = 2$, y así sucesivamente.]

20. Tres parejas y dos individuos solteros han sido invitados a un seminario de inversión y han aceptado asistir. Suponga que la probabilidad de que cualquier pareja o individuo particular llegue tarde es de .4 (una pareja viajará en el mismo vehículo, así que ambos llegarán a tiempo o bien ambos llegarán tarde). Suponga que diferentes parejas e individuos llegan puntuales o tarde independientemente unos de otros. Sea X = el número de personas que llegan tarde al seminario.

a. Determine la función de masa de probabilidad de X. [*Sugerencia*: designe las tres parejas #1, #2 y #3 y los dos individuos #4 y #5.]

b. Obtenga la función de distribución acumulativa de X y úsela para calcular $P(2 \leq X \leq 6)$.

21. Suponga que lee los números de este año del *New York Times* y que anota cada número que aparece en un artículo de noticias: el ingreso de un oficial ejecutivo en jefe, el número de cajas de vino producidas por una compañía vinícola, la contribución caritativa total de un político durante el año fiscal previo, la edad de una celebridad, y así sucesivamente. Ahora enfóquese en el primer dígito de cada número, el cual podría ser 1, 2, . . . , 8 o 9. Su primer pensamiento podría que el primer dígito X de un número seleccionado al azar sería igualmente probable que fuera una de las nueve posibilidades (una distribución uniforme discreta). Sin embargo, mucha evidencia empírica así como también algunos argumentos teóricos sugieren una distribución de probabilidad alternativa llamada *ley de Benford*:

$$p(x) = P(\text{el primer dígito es } x) = \log_{10}\left(\frac{x+1}{x}\right) \quad x = 1, 2, \ldots, 9$$

a. Sin calcular probabilidades individuales de esta fórmula, demuestre que especifica una función de masa de probabilidad legítima.

b. Ahora calcule las probabilidades individuales y compare con la distribución uniforme discreta correspondiente.

c. Obtenga la función de distribución acumulativa de X.

d. Utilizando la función de distribución acumulativa, ¿cuál es la probabilidad de que el primer dígito sea cuando mucho 3? ¿Por lo menos 5?

[*Nota*: la ley de Benford es la base de algunos procedimientos de auditoría utilizados para detectar fraudes en reportes financieros, por ejemplo, por el Servicio de Ingresos Internos.]

22. Remítase al ejercicio 13 y calcule y trace la gráfica de la función de distribución acumulativa $F(x)$. Luego utilícela para calcular las probabilidades de los eventos dados en los incisos (a)–(d) de dicho problema.

23. Una organización de protección al consumidor que habitualmente evalúa automóviles nuevos reporta el número de defectos importantes encontrados en cada carro examinado. Sea X el número de defectos importantes en un carro seleccionado al azar de cierto tipo. La función de distribución acumulativa de X es la siguiente:

$$F(x) = \begin{cases} 0 & x < 0 \\ .06 & 0 \le x < 1 \\ .19 & 1 \le x < 2 \\ .39 & 2 \le x < 3 \\ .67 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ .97 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

Calcule las siguientes probabilidades directamente con la función de distribución acumulativa:

a. $p(2)$, es decir, $P(X = 2)$ **b.** $P(X > 3)$
c. $P(2 \le X \le 5)$ **d.** $P(2 < X < 5)$

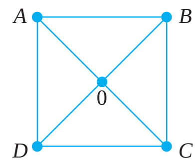
24. Una compañía de seguros ofrece a sus asegurados varias opciones diferentes de pago de primas. Para un asegurado seleccionado al azar, sea $X =$ el número de meses entre pagos sucesivos. La función de distribución acumulativa es la siguiente:

$$F(x) = \begin{cases} 0 & x < 1 \\ .30 & 1 \le x < 3 \\ .40 & 3 \le x < 4 \\ .45 & 4 \le x < 6 \\ .60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

a. ¿Cuál es la función de masa de probabilidad de X?
b. Con sólo la función de distribución acumulativa, calcule $P(3 \le X \le 6)$ y $P(4 \le X)$.

25. En el ejemplo 3.12, sea $Y =$ el número de niñas nacidas antes de que termine el experimento. Con $p = P(B)$ y $1 - p = P(G)$, ¿cuál es la función de masa de probabilidad de Y? [*Sugerencia*: primero ponga en lista los posibles valores de Y, inicie con el más pequeño y continúe hasta que encuentre una fórmula general.]

26. Alvie Singer vive en 0 en el diagrama adjunto y sus cuatro amigos viven en A, B, C y D. Un día Alvie decide visitarlos, así que lanza al aire una moneda imparcial dos veces para decidir a cuál de los cuatro visitar. Una vez que está en la casa de uno de sus amigos, regresará a su casa o bien proseguirá a una de las dos casas adyacentes (tales como 0, A o C, cuando está en B) con cada una de las tres posibilidades cuya probabilidad es $\frac{1}{3}$. De este modo, Alvie continúa visitando a sus amigos hasta que regresa a casa.



a. Sea $X =$ el número de veces que Alvie visita a un amigo. Obtenga la función de masa de probabilidad de X.
b. Sea $Y =$ el número de segmentos de línea recta que Alvie recorre (incluidos los que conducen a 0 o que parten de ahí). ¿Cuál es la función de masa de probabilidad de Y?
c. Suponga que sus amigas viven en A y C y sus amigos en B y D. Si $Z =$ el número de visitas a amigas, ¿cuál es la función de masa de probabilidad de Z?

27. Después de que todos los estudiantes salieron del salón de clases, un profesor de estadística nota que cuatro ejemplares del texto se quedaron debajo de los escritorios. Al principio de la siguiente clase, el profesor distribuye los cuatro libros al azar a cada uno de los cuatro estudiantes (1, 2, 3 y 4) que dicen haber dejado los libros. Un posible resultado es que 1 reciba el libro de 2, que 2 reciba el libro de 4 y que 3 reciba su propio libro y que 4 reciba el libro de 1. Este resultado puede ser abreviado como (2, 4, 3, 1).
a. Mencione los otros 23 resultados posibles.
b. Si X es el número de estudiantes que reciben su propio libro, determine la función de masa de probabilidad de X.

28. Demuestre que la función de distribución acumulativa de $F(x)$ es no decreciente; es decir, $x_1 < x_2$ implica que $F(x_1) \le F(x_2)$. ¿En qué condición será $F(x_1) = F(x_2)$?

3.3 Valores esperados

Considérese una universidad que tiene 15,000 estudiantes y sea $X =$ el número de cursos en los cuales está inscrito un estudiante seleccionado al azar. La función de masa de probabilidad de X se determina como sigue. Como $p(1) = .01$, se sabe que $(.01) \cdot (15{,}000) = 150$ de los estudiantes están inscritos en un curso y asimismo con los demás valores de x.

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02
Número de inscritos	150	450	1950	3750	5850	2550	300

(3.6)

El número promedio de cursos por estudiante o el valor promedio de X en la población se obtiene al calcular el número total de cursos tomados por todos los estudiantes y dividir entre el número total de estudiantes. Como cada uno de los 150 estudiantes está tomando un curso, estos 150 contribuyen con 150 cursos al total. Asimismo, 450 estudiantes contribuyen con 2(450) cursos, y así sucesivamente. El valor promedio de la población de X es entonces

$$\frac{1(150) + 2(450) + 3(1950) + \cdots + 7(300)}{15{,}000} = 4.57 \tag{3.7}$$

Como $150/15{,}000 = .01 = p(1)$, $450/15{,}000 = .03 = p(2)$, y así sucesivamente, una expresión alternativa para (3.7) es

$$1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \tag{3.8}$$

La expresión (3.8) muestra que para calcular el valor promedio de la población de X, sólo se necesitan los valores posibles de X junto con sus probabilidades (proporciones). En particular, el tamaño de la población no viene al caso en tanto la función de masa de probabilidad esté dada por (3.6). El valor promedio o medio de X es entonces el promedio *ponderado* de los posibles valores 1, . . . , 7, donde las ponderaciones son las probabilidades de esos valores.

Valor esperado de *X*

DEFINICIÓN

Sea X una variable aleatoria discreta con un conjunto de valores posibles D y una función de masa de probabilidad $p(x)$. El **valor esperado** o **valor medio** de X, denotado por $E(X)$ o μ_X o sólo μ, es

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Ejemplo 3.16 Para la función de masa de probabilidad de X = *número de cursos* en (3.6),

$$\begin{aligned} \mu &= 1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \\ &= (1)(.01) + 2(.03) + \cdots + (7)(.02) \\ &= .01 + .06 + .39 + 1.00 + 1.95 + 1.02 + .14 = 4.57 \end{aligned}$$

Si se piensa en la población como compuesta de los valores 1, 2, . . . , 7 de X, entonces $\mu = 4.57$ es la media de la población. En lo que sigue, a menudo se hará referencia a μ como la *media de la población* en lugar de la media de X en la población. Tenga en cuenta que μ aquí no es 4, el promedio normal de 1, . . . , 7, porque la distribución pone más peso en 4, 5 y 6 que en otros valores de X. ■

En el ejemplo 3.16, el valor esperado μ fue 4.57, el cual no es un valor posible de X. La palabra *esperado* deberá interpretarse con precaución porque no se esperaría ver un valor de X de 4.57 cuando se selecciona un solo estudiante.

Ejemplo 3.17 Exactamente después de nacer, cada recién nacido es evaluado en una escala llamada escala de Apgar. Las evaluaciones posibles son 0, 1, . . . , 10, con la evaluación del niño determinada por color, tono muscular, esfuerzo para respirar, ritmo cardiaco e irritabilidad refleja (la mejor evaluación posible es 10). Sea X la evaluación Apgar de un niño seleccionado al azar nacido en cierto hospital durante el siguiente año y supóngase que la función de masa de probabilidad de X es

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

Entonces el valor medio de X es

$$\begin{aligned} E(X) = \mu &= 0(.002) + 1(.001) + 2(.002) \\ &\quad + \cdots + 8(.25) + 9(.12) + 10(.01) \\ &= 7.15 \end{aligned}$$

De nuevo, μ no es un valor posible de la variable X. Además, como la variable se refiere a un niño futuro, no existe ninguna población concreta a la cual se podría referir μ. En cambio, la función de masa de probabilidad se considera como un modelo de una población conceptual compuesta de los valores 0, 1, 2, . . . , 10. El valor medio de esta población conceptual es entonces $\mu = 7.15$. ■

Ejemplo 3.18 Sea $X = 1$ si un vehículo seleccionado al azar aprueba un diagnóstico de emisiones y $X = 0$ si no. Entonces X es una variable aleatoria de Bernoulli con función de masa de probabilidad $p(1) = p$ y $p(0) = 1 - p$ a partir de la cual $E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1 - p) + 1(p) = p$. Es decir, el valor esperado de X es exactamente la probabilidad de que X tome el valor de 1. Si se conceptualiza una población compuesta de ceros en la proporción $1 - p$ y números 1 en la proporción p, entonces el promedio de la población es $\mu = p$. ■

Ejemplo 3.19 La forma general de la función de masa de probabilidad de X = número de niños nacidos hasta el primer varón incluido éste

$$p(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

De acuerdo con la definición,

$$E(X) = \sum_D x \cdot p(x) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = p\sum_{x=1}^{\infty}\left[-\frac{d}{dp}(1-p)^x\right] \tag{3.9}$$

Si se intercambia el orden en que se evalúan la derivada y la suma, la suma es la de una serie geométrica. Una vez que se calcula la suma, se saca la derivada y el resultado final es $E(X) = 1/p$. Si p se aproxima a 1, se espera ver que nazca un varón muy pronto, mientras que si p se aproxima a 0, se esperan muchos nacimientos antes del primer varón. Con $p = .5$, $E(X) = 2$. ■

Existe otra interpretación frecuentemente utilizada de μ. Considere la posibilidad de observar un primer valor x_1 de X, a continuación, un segundo valor x_2, un tercer valor x_3 y así sucesivamente. Después de hacer esto un gran número de veces, se calcula el promedio de la muestra de las x_i observadas. Este promedio será típicamente muy cercano a μ. Es decir, μ puede interpretarse como el promedio a largo plazo del valor observado de X cuando el experimento se realiza en varias ocasiones. En el ejemplo 3.17, el promedio a largo plazo de Apgar es $\mu = 7.15$.

Ejemplo 3.20 Sea X el número de entrevistas que un estudiante sostiene antes de conseguir un trabajo, y tiene la función de masa de probabilidad

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

donde k se elige de modo que $\sum_{x=1}^{\infty}(k/x^2) = 1$. (En un curso de matemáticas de series infinitas se demuestra que $\sum_{x=1}^{\infty}(1/x^2) < \infty$, lo cual implica que tal k existe, pero su valor exacto no interesa.) El valor esperado de X es

$$\mu = E(X) = \sum_{x=1}^{\infty} x \cdot \frac{k}{x^2} = k\sum_{x=1}^{\infty}\frac{1}{x} \tag{3.10}$$

La suma del lado derecho de la ecuación (3.10) es la famosa serie armónica de matemáticas y se puede demostrar que es igual a ∞. $E(X)$ no es finita en este caso porque $p(x)$ no disminuye suficientemente rápido a medida que x se incrementa; los estadísticos dicen que la distribución de probabilidad de X tiene "una cola gruesa". Si se selecciona una secuencia de valores X utilizando esta distribución, el promedio muestral no se establecerá en un número finito sino que tenderá a crecer sin límite.

Los estadísticos utilizan la frase "colas gruesas" en conexión con cualquier distribución con una gran cantidad de probabilidad alejada de μ (así que las colas gruesas no requieren $\mu = \infty$). Tales colas gruesas hacen difícil hacer inferencias sobre μ. ■

Valor esperado de una función

A menudo interesará poner atención al valor esperado de alguna función $h(X)$ en lugar de sólo en $E(X)$.

Ejemplo 3.21 Suponga que una librería adquiere diez ejemplares de un libro a \$6.00 cada uno para venderlos a \$12.00 en el entendimiento de que al final de un periodo de 3 meses cualquier ejemplar no vendido puede ser compensado por \$2.00. Si $X =$ el número de ejemplares vendidos, entonces el ingreso neto $= h(X) = 12X + 2(10 - X) - 60 = 10X - 40$. ¿Cuál es entonces el ingreso neto esperado? ■

El siguiente ejemplo sugiere una forma fácil de calcular el valor esperado de $h(X)$.

Ejemplo 3.22 El costo de cierta prueba de diagnóstico de un vehículo depende del número de cilindros X en el motor. Supóngase que la función de costo está dada por $h(X) = 20 + 3X + .5X^2$. Como X es una variable aleatoria, también lo es $Y = h(X)$. Las funciones de masa de probabilidad de X y Y son las siguientes

x	4	6	8
$p(x)$	.5	.3	.2

$\Rightarrow$

y	40	56	76
$p(y)$	.5	.3	.2

Con D^* denotando posibles valores de Y,

$$\begin{aligned} E(Y) = E[h(X)] &= \sum_{D^*} y \cdot p(y) \\ &= (40)(.5) + (56)(.3) + (76)(.2) \\ &= h(4) \cdot (.5) + h(6) \cdot (.3) + h(8) \cdot (.2) \\ &= \sum_{D} h(x) \cdot p(x) \end{aligned} \tag{3.11}$$

De acuerdo con la ecuación (3.11), no fue necesario determinar la función de masa de probabilidad de Y para obtener $E(Y)$; en su lugar, el valor esperado deseado es un promedio ponderado de los posibles valores de $h(x)$ (y no de x). ■

PROPOSICIÓN

Si la variable aleatoria X tiene un conjunto de posibles valores D y una función de masa de probabilidad $p(x)$, entonces el valor esperado de cualquier función $h(X)$, denotada por $E[h(X)]$ o $\mu_{h(X)}$, se calcula con

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

Esto es, $E[h(X)]$ se calcula del mismo modo que $E(X)$, excepto que $h(x)$ sustituye a x.

Ejemplo 3.23 Una tienda de computadoras adquirió tres computadoras de un tipo a \$500 cada una. Las venderá a \$1000 cada una. El fabricante se comprometió a readquirir cualquier computadora que no se haya vendido después de un periodo especificado a \$200 cada una. Sea X el número de computadoras vendidas y suponga que $p(0) = .1$, $p(1) = .2$, $p(2) = .3$ y $p(3) = .4$. Con $h(X)$ denotando la utilidad asociada con la venta de X unidades, la información dada implica que $h(X) = \text{ingreso} - \text{costo} = 1000X + 200(3 - X) - 1500 = 800X - 900$. La utilidad esperada es entonces

$$\begin{aligned} E[h(X)] &= h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) \\ &= (-900)(.1) + (-100)(.2) + (700)(.3) + (1500)(.4) \\ &= \$700 \end{aligned}$$

■

Reglas de valor esperado

La función de interés $h(X)$ es con bastante frecuencia una función lineal $aX + b$. En este caso, $E[h(X)]$ es fácil de calcular a partir de $E(X)$.

PROPOSICIÓN

$$E(aX + b) = a \cdot E(X) + b$$

(O, con notación alternativa, $\mu_{aX+b} = a \cdot \mu_X + b$)

Parafraseando, el valor esperado de una función lineal es igual a la función lineal evaluada con el valor esperado $E(X)$. Como $h(X)$ en el ejemplo 3.23 es lineal y $E(X) = 2$, $E[h(x)] = 800(2) - 900 = \$700$, como antes.

Comprobación

$$\begin{aligned} E(aX + b) &= \sum_D (ax + b) \cdot p(x) = a\sum_D x \cdot p(x) + b\sum_D p(x) \\ &= aE(X) + b \end{aligned}$$

■

Dos casos especiales de proposición producen dos reglas importantes de valor esperado.

1. Con cualquier constante a, $E(aX) = a \cdot E(X)$ (considérese $b = 0$). (3.12)
2. Con cualquier constante b, $E(X + b) = E(X) + b$ (considérese $a = 1$).

La multiplicación de X por una constante a cambia por lo general la unidad de medición; por ejemplo, de pulgadas a centímetros, donde $a = 2.54$, etc.). La regla 1 dice que el valor esperado en las nuevas unidades es igual al valor esperado en las viejas unidades multiplicado por el factor de conversión a. Asimismo, si se agrega una constante b a cada valor posible de X, entonces el valor esperado se desplazará en esa misma cantidad constante.

Varianza de X

El valor esperado de X describe dónde está centrada la distribución de probabilidad. Utilizando la analogía física de colocar una masa puntual $p(x)$ en el valor x sobre un eje unidimensional que estuviera soportado por un fulcro colocado en μ, el eje no tendería a ladearse. Esto se ilustra para dos distribuciones diferentes en la figura 3.7.

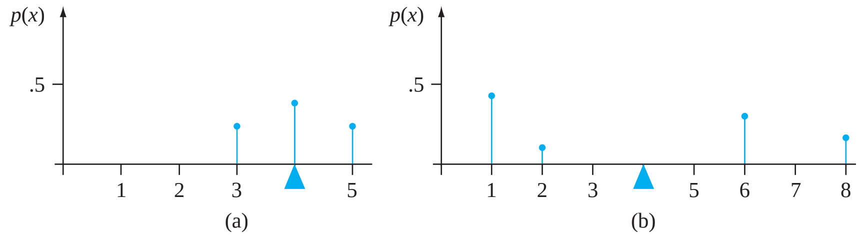


Figura 3.7 Dos diferentes distribuciones de probabilidad con $\mu = 4$

Aunque ambas distribuciones ilustradas en la figura 3.7 tienen el mismo centro μ, la distribución de la figura 3.7(b) tiene una mayor dispersión o variabilidad que la de la figura 3.7(a). Se utilizará la varianza de X para evaluar la cantidad de variabilidad en (la distribución de) X, del mismo modo que se utilizó s^2 en el capítulo 1 para medir la variabilidad en una muestra.

DEFINICIÓN

Sea $p(x)$ la función de masa de probabilidad de X y μ su valor esperado. En ese caso la **varianza** de X, denotada por $V(X)$ o σ_X^2, o simplemente σ^2, es

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es

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

La cantidad $h(X) = (X - \mu)^2$ es la desviación al cuadrado de X con respecto a su media y σ^2 es la desviación al cuadrado esperada, es decir, el promedio ponderado de las desviaciones al cuadrado, donde las ponderaciones son probabilidades de la distribución. Si la mayor parte de la distribución de probabilidad está cerca de μ, entonces σ^2 será relativamente pequeña. Sin embargo, si existen valores x alejados de μ que tienen una gran $p(x)$, en ese caso σ^2 será bastante grande. A grandes rasgos, σ se puede interpretar como el tamaño de una desviación representativa del valor medio μ. Así que si $\sigma = 10$, entonces en una larga secuencia de valores observados X, algunos se apartarán de μ por más de 10, mientras que otros estarán más cerca de la media que eso, una desviación típica de la media será del orden de 10.

Ejemplo 3.24 Una biblioteca tiene un límite superior de 6 en el número de videos que puede sacar una persona a la vez. Tenga en cuenta sólo a quienes echan un vistazo a los videos y deje que X denote el número de videos que pide prestados para una persona seleccionada al azar. La función de masa de probabilidad de X es la siguiente:

x	1	2	3	4	5	6
$p(x)$	.30	.25	.15	.05	.10	.15

Es fácil ver que el valor esperado de X es $\mu = 2.85$. La varianza de X es entonces

$$\begin{aligned} V(X) = \sigma^2 &= \sum_{x=1}^{6} (x - 2.85)^2 \cdot p(x) \\ &= (1 - 2.85)^2(.30) + (2 - 2.85)^2(.25) + \cdots + (6 - 2.85)^2(.15) = 3.2275 \end{aligned}$$

La desviación estándar de X es $\sigma = \sqrt{3.2275} = 1.800$. ■

Cuando la función de masa de probabilidad $p(x)$ especifica un modelo matemático para la distribución de los valores de la población, tanto σ^2 como σ miden la dispersión de los valores en la población; σ^2 es la varianza de la población y σ es su desviación estándar.

Fórmula abreviada para σ^2

El número de operaciones aritméticas necesarias para calcular σ^2 puede reducirse si se utiliza una fórmula alternativa.

PROPOSICIÓN

$$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x)\right] - \mu^2 = E(X^2) - [E(X)]^2$$

Al utilizar esta fórmula, $E(X^2)$ se calcula primero sin ninguna sustracción; acto seguido $E(X)$ se calcula, se eleva al cuadrado y se resta (una vez) de $E(X^2)$.

Ejemplo 3.25 (Continuación del ejemplo 3.24)

La función de masa de probabilidad del número X de videos prestados se dio como $p(1) = .30$, $p(2) = .25$, $p(3) = .15$, $p(4) = .05$, $p(5) = .10$ y $p(6) = .15$, a partir de las cuales $\mu = 2.85$ y

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot p(x) = (1^2)(.30) + (2^2)(.25) + \cdots + (6^2)(.15) = 11.35$$

Por lo tanto, $\sigma^2 = 11.35 - (2.85)^2 = 3.2275$, como se obtuvo previamente de la definición. ■

Demostración de la fórmula abreviada Desarróllese $(x - \mu)^2$ en la definición de σ^2 para obtener $x^2 - 2\mu x + \mu^2$, y luego lleve $\sum$ a cada uno de los tres términos:

$$\begin{aligned}\sigma^2 &= \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2\end{aligned}$$ ■

Reglas de varianza

La varianza de $h(X)$ es el valor esperado de la diferencia al cuadrado entre $h(X)$ y su valor esperado:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x) \tag{3.13}$$

Cuando $h(X) = aX + b$, una función lineal,

$$h(x) - E[h(X)] = ax + b - (a\mu + b) = a(x - \mu)$$

Sustituyendo esto en la ecuación (3.13) se obtiene una relación simple entre $V[h(X)]$ y $V(X)$:

PROPOSICIÓN

$$V(aX + b) = \sigma^2_{aX+b} = a^2 \cdot \sigma^2_X \text{ y } \sigma_{aX+b} = |a| \cdot \sigma_x$$

En particular,

$$\sigma_{aX} = |a| \cdot \sigma_X, \quad \sigma_{X+b} = \sigma_X \tag{3.14}$$

El valor absoluto es necesario porque a podría ser negativa, no obstante una desviación estándar no puede serlo. Casi siempre la multiplicación por a corresponde a un cambio de la unidad de medición (p. ej., kg a lb o dólares a euros). De acuerdo con la primera relación en (3.14), la desviación estándar en la nueva unidad es la desviación estándar original multiplicada por el factor de conversión. La segunda relación dice que la adición o sustracción de una constante no impacta la variabilidad; simplemente desplaza la distribución a la derecha o izquierda.

Ejemplo 3.26 En el problema de ventas de computadoras del ejemplo 3.23, $E(X) = 2$ y

$$E(X^2) = (0)^2(.1) + (1)^2(.2) + (2)^2(.3) + (3)^2(.4) = 5$$

así que $V(X) = 5 - (2)^2 = 1$. La función de utilidad $h(X) = 800X - 900$ tiene entonces la varianza $(800)^2 \cdot V(X) = (640{,}000)(1) = 640{,}000$ y la desviación estándar 800. ■

EJERCICIOS Sección 3.3 (29–45)

29. La función de masa de probabilidad de la cantidad de memoria X (GB) en una unidad flash comprada fue dada en el ejemplo 3.13 como

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

Calcule lo siguiente:
a. $E(X)$
b. $V(X)$ directamente a partir de la definición
c. La desviación estándar de X
d. $V(X)$ por medio de la fórmula abreviada

30. Se selecciona al azar un individuo que tiene asegurado su automóvil con una compañía. Sea Y el número de infracciones de tránsito por las que el individuo fue citado durante los últimos tres años. La función de masa de probabilidad de Y es

y	0	1	2	3
$p(y)$	.60	.25	.10	.05

a. Calcule $E(Y)$.
b. Suponga que un individuo con Y infracciones incurre en un recargo de $\$100Y^2$. Calcule el monto esperado del recargo.

31. Remítase al ejercicio 12 y calcule $V(Y)$ y σ_Y. Determine entonces la probabilidad de que Y esté dentro de una desviación estándar de 1 de su valor medio.

32. Un distribuidor de enseres para el hogar vende tres modelos de congeladores verticales de 13.5, 15.9 y 19.1 pies cúbicos de espacio de almacenamiento, respectivamente. Sea X = la cantidad de espacio de almacenamiento adquirido por el siguiente cliente que compre un congelador. Suponga que X tiene la función de masa de probabilidad

x	13.5	15.9	19.1
$p(x)$	.2	.5	.3

a. Calcule $E(X)$, $E(X^2)$ y $V(X)$.
b. Si el precio de un congelador de X pies cúbicos de capacidad es $25X - 8.5$, ¿cuál es el precio esperado pagado por el siguiente cliente que compre un congelador?
c. ¿Cuál es la varianza del precio $25X - 8.5$ pagado por el siguiente cliente?
d. Suponga que aunque la capacidad nominal de un congelador es X, la real es $h(X) = X - .01X^2$. ¿Cuál es la capacidad real esperada del congelador adquirido por el siguiente cliente?

33. Sea X una variable aleatoria de Bernoulli con función de masa de probabilidad como en el ejemplo 3.18.
a. Calcule $E(X^2)$
b. Demuestre que $V(X) = p(1 - p)$.
c. Calcule $E(X^{79})$.

34. Suponga que el número de plantas de un tipo particular encontradas en una región rectangular de muestreo (llamada cuadrado por los ecologistas) en cierta área geográfica es una variable aleatoria X con función de masa de probabilidad

$$p(x) = \begin{cases} c/x^3 & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

¿Es $E(X)$ finita? Justifique su respuesta (ésta es otra distribución que los estadísticos llamarían de cola gruesa).

35. Un pequeño mercado ordena ejemplares de cierta revista para su exhibidor de revistas cada semana. Sea X = demanda de la revista, con función de masa de probabilidad

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suponga que el propietario de la tienda paga \$2.00 por cada ejemplar de la revista y el precio para los consumidores es de \$4.00. Si las revistas que se quedan al final de la semana no tienen valor de recuperación, ¿es mejor ordenar tres o cuatro ejemplares de la revista? [*Sugerencia*: para tres o cuatro ejemplares ordenados, exprese el ingreso neto como una función de la demanda X, y luego calcule el ingreso esperado.]

36. Sea X el daño incurrido (en dólares) en un tipo de accidente durante un año dado. Valores posibles de X son 0, 1000, 5000 y 10,000, con probabilidades de .8, .1, .08 y .02, respectivamente. Una compañía particular ofrece una póliza con deducible de \$500. Si la compañía desea que su utilidad esperada sea de \$100, ¿qué cantidad de prima deberá cobrar?

37. Los n candidatos para un trabajo fueron clasificados como 1, 2, 3, . . . , n. Sea X = la clasificación de un candidato seleccionado al azar, de modo que X tenga la función de masa de probabilidad

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \dots, n \\ 0 & \text{de lo contrario} \end{cases}$$

(ésta se llama *distribución uniforme discreta*). Calcule $E(X)$ y $V(X)$ por medio de la fórmula abreviada. [*Sugerencia*: la suma de los primeros n enteros positivos es $n(n + 1)/2$, mientras que la suma de sus cuadrados es $n(n + 1)(2n + 1)/6$.]

38. Sea X = el resultado cuando un dado imparcial es lanzado una vez. Si antes de lanzar el dado le ofrecen (1/3.5) dólares o $h(X) = 1/X$ dólares, ¿aceptaría la suma garantizada o jugaría? [*Nota*: generalmente no es cierto que $1/E(X) = E(1/X)$.]

39. Una compañía de productos químicos en la actualidad tiene en existencia 100 lb de un producto químico, el cual se vende a sus clientes en lotes de 5 lb. Sea X = el número de lotes solicitados por un cliente seleccionado al azar y suponga que X tiene la función de masa de probabilidad

x	1	2	3	4
$p(x)$	.2	.4	.3	.1

Calcule $E(X)$ y $V(X)$. Calcule en seguida el número esperado de libras que quedan una vez que se envía el pedido del siguiente cliente y la varianza del número de libras que quedan. [*Sugerencia*: el número de libras que quedan es una función lineal de X.]

40. **a.** Trace una gráfica lineal de la función de masa de probabilidad de X en el ejercicio 35. En seguida determine la función de masa de probabilidad de $-X$ y trace su gráfica lineal. Con base en estas dos figuras, ¿qué se puede decir sobre $V(X)$ y $V(-X)$?

b. Use la proposición que implica $V(aX + b)$ para establecer una relación general entre $V(X)$ y $V(-X)$.

41. Use la definición en la expresión (3.13) para comprobar que $V(aX + b) = a^2 \cdot \sigma_X^2$. [*Sugerencia*: con $h(X) = aX + b$, $E[h(X)] = a\mu + b$, donde $\mu = E(X)$.]

42. Suponga $E(X) = 5$ y $E[X(X - 1)] = 27.5$. ¿Cuál es

a. $E(X^2)$? [*Sugerencia*: $E[X(X - 1)] = E(X^2 - X] = E(X^2) - E(X)$]?

b. $V(X)$?

c. la relación general entre las cantidades $E(X)$, $E[X(X - 1)]$ y $V(X)$?

43. Escriba una regla general para $E(X - c)$, donde c es una constante. ¿Qué sucede cuando hace $c = \mu$, el valor esperado de X?

44. Un resultado llamado **desigualdad de Chebyshev** establece que para cualquier distribución de probabilidad de una variable aleatoria X y cualquier número k que por lo menos sea 1, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$. En palabras, la posibilidad de que el valor de X quede por lo menos a k desviaciones estándar de su media es cuando mucho $1/k^2$.

a. ¿Cuál es el valor del límite superior con $k = 2$? ¿$k = 3$? ¿$k = 4$? ¿$k = 5$? ¿$k = 10$?

b. Calcule μ y σ para la distribución del ejercicio 13. Evalúe en seguida $P(|X - \mu| \geq k\sigma)$ con los valores de k dados en el inciso (a). ¿Qué sugiere esto sobre el límite superior con respecto a la probabilidad correspondiente?

c. Sea X con los valores posibles -1, 0 y 1, con las probabilidades $\frac{1}{18}$, $\frac{8}{9}$ y $\frac{1}{18}$, respectivamente. ¿Cuál es $P(|X - \mu| \geq 3\sigma)$ y cómo se compara con el límite correspondiente?

d. Dé una distribución para la cual $P(|X - \mu| \geq 5\sigma) = .04$.

45. Si $a \leq X \leq b$, demuestre que $a \leq E(X) \leq b$.

3.4 Distribución de probabilidad binomial

Existen muchos experimentos que se ajustan exacta o aproximadamente a la siguiente lista de requerimientos.

1. El experimento consta de una secuencia de n experimentos más pequeños llamados *ensayos*, donde n se fija antes del experimento.
2. Cada ensayo puede dar por resultado uno de los mismos dos resultados posibles (ensayos dicotómicos), los cuales se denotan como éxito (S) y falla (F).
3. Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.
4. La probabilidad de éxito $P(S)$ es constante de un ensayo a otro; esta probabilidad se denota por p.

DEFINICIÓN

Un experimento para el que se satisfacen las condiciones 1–4 se llama **experimento binomial.**

Ejemplo 3.27 La misma moneda se lanza al aire sucesiva e independientemente n veces. De manera arbitraria se utiliza S para denotar el resultado H (caras) y F para denotar el resultado T (cruces). Entonces este experimento satisface las condiciones 1–4. El lanzamiento al aire de una tachuela n veces, con S = punta hacia arriba y F = punta hacia abajo, también da por resultado un experimento binomial. ■

Muchos experimentos implican una secuencia de ensayos independientes para los cuales existen más de dos resultados posibles en cualquier ensayo. Entonces, un experimento binomial puede crearse dividiendo los posibles resultados en dos grupos.

Ejemplo 3.28 El color de las semillas de chícharo lo determina un solo locus genético. Si los dos alelos en este locus genético son AA o Aa (el genotipo), entonces el chícharo será amarillo (el fenotipo) y si el alelo es aa, el chícharo será verde. Suponga que se aparean 20 semillas Aa y se cruzan las dos semillas en cada uno de los diez pares para obtener diez nuevos genotipos. Designe a cuada nuevo genotipo como éxito (S) si es aa y falla (F) si es lo contrario. Entonces con esta identificación de S y F, el experimento es binomial con $n = 10$ y $p = P$ (genotipo aa). Si es igualmente probable que cada miembro del par contribuya con a o A, entonces $p = P(a) \cdot P(a) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$. ■

Ejemplo 3.29 Suponga que una ciudad tiene 50 restaurantes autorizados, de los cuales 15 han cometido en la actualidad una seria violación del código sanitario y los otros 35 no han cometido violaciones serias. Hay cinco inspectores, cada uno de los cuales inspeccionará un restaurante durante la semana entrante. El nombre de cada restaurante se anota en un pedacito de papel diferente y a continuación se mezclan perfectamente, cada inspector a su vez saca uno de los papelitos *sin reemplazarlos*. Anótese el ensayo i-ésimo como éxito si el restaurante i-ésimo seleccionado ($i = 1, \ldots, 5$) no ha cometido violaciones serias. Entonces

$$P(S \text{ en el primer ensayo}) = \frac{35}{50} = .70$$

y

$$\begin{aligned} P(S \text{ en el segundo ensayo}) &= P(SS) + P(FS) \\ &= P(\text{segundo } S \mid \text{primer } S)P(\text{primer } S) \\ &\quad + P(\text{segundo } S \mid \text{primera } F)P(\text{primera } F) \\ &= \frac{34}{49}\cdot\frac{35}{50} + \frac{35}{49}\cdot\frac{15}{50} = \frac{35}{50}\left(\frac{34}{49}+\frac{15}{49}\right) = \frac{35}{50} = .70 \end{aligned}$$

De manera similar, se puede demostrar que $P(S$ en el ensayo i-ésimo$) = .70$ con $i = 3, 4, 5$. Sin embargo,

$$P(S \text{ en el quinto ensayo} \mid SSSS) = \frac{31}{46} = .67$$

mientras que

$$P(S \text{ en el quinto ensayo} \mid FFFF) = \frac{35}{46} = .76$$

El experimento no es binomial porque los ensayos no son independientes. En general, si se muestrea sin reemplazo, el experimento no producirá ensayos independientes. Si cada papelito hubiera sido reemplazado después de ser sacado, entonces los ensayos habrían sido independientes, pero esto podría haber dado por resultado que el mismo restaurante fuera inspeccionado por más de un inspector. ■

Ejemplo 3.30 Un estado tiene 500,000 conductores con licencia, de los cuales 400,000 están asegurados. Se selecciona una muestra de 10 conductores sin reemplazo. El ensayo *i*-ésimo se denota *S* si el conductor *i*ésimo seleccionado está asegurado. Aunque esta situación parecería idéntica a la del ejemplo 3.29, la diferencia importante es que el tamaño de la población muestreada es muy grande con respecto al tamaño de la muestra. En este caso

$$P(S \text{ en } 2 \mid S \text{ en } 1) = \frac{399{,}999}{499{,}999} = .80000$$

y

$$P(S \text{ en } 10 \mid S \text{ en los primeros } 9) = \frac{399{,}991}{499{,}991} = .799996 \approx .80000$$

Estos cálculos sugieren que aunque los ensayos no son exactamente independientes, las probabilidades condicionales difieren tan poco una de otra que para propósitos prácticos los ensayos se consideran independientes con la constante $P(S) = .8$. Por lo tanto, para una muy buena aproximación, el experimento es binomial con $n = 10$ y $p = .8$. ■

Se utilizará la siguiente regla empírica para decidir si un experimento "sin reemplazo" puede ser tratado como un experimento binomial.

REGLA

Considérese muestreo sin reemplazo de una población dicotómica de tamaño N. Si el tamaño de la muestra (número de ensayos) n es cuando mucho 5% del tamaño de la población, el experimento puede ser analizado como si fuera exactamente un experimento binomial.

Por "analizado" se quiere decir que las probabilidades basadas en suposiciones de experimento binomial se aproximarán bastante a las probabilidades reales "sin reemplazo", las que generalmente son más difíciles de calcular. En el ejemplo 3.29, $n/N = 5/50 = .1 > .05$, de modo que el experimento binomial no es una buena aproximación, pero en el ejemplo 3.30, $n/N = 10/500{,}000 < .05$.

Variable y distribución aleatoria binomial

En la mayoría de los experimentos binomiales, lo que interesa es el número total de los éxitos (S), en lugar del conocimiento de qué ensayos dieron los éxitos.

DEFINICIÓN

La **variable aleatoria binomial** $\boldsymbol{X}$ asociada con un experimento binomial que consiste en n ensayos se define como

$$X = \text{el número de los } S \text{ entre los } n \text{ ensayos}$$

Supóngase, por ejemplo, que $n = 3$. Entonces existen ocho posibles resultados para el experimento:

$$SSS \quad SSF \quad SFS \quad SFF \quad FSS \quad FSF \quad FFS \quad FFF$$

Por la definición de X, $X(SSF) = 2$, $X(SFF) = 1$, y así sucesivamente. Valores posibles de X en un experimento de n ensayos son $x = 0, 1, 2, \ldots, n$. A menudo se escribirá $X \sim \text{Bin}(n, p)$ para indicar que X es una variable aleatoria binomial basada en n ensayos con probabilidad de éxito p.

NOTACIÓN

Como la función de masa de probabilidad de una variable aleatoria binomial X depende de los dos parámetros n y p, la función de masa de probabilidad se denota por $b(x; n, p)$.

Considérese primero el caso $n = 4$ para el cual cada resultado, su probabilidad y su valor x correspondiente se dan en la tabla 3.1. Por ejemplo,

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) \quad \text{(ensayos independientes)} \\ &= p \cdot p \cdot (1 - p) \cdot p \quad \text{(constante } P(S)\text{)} \\ &= p^3 \cdot (1 - p) \end{aligned}$$

Tabla 3.1 Resultados y probabilidades para un experimento binomial con cuatro intentos

Resultado	x	Probabilidad	Resultado	x	Probabilidad
SSSS	4	p^4	*FSSS*	3	$p^3(1 - p)$
SSSF	3	$p^3(1 - p)$	*FSSF*	2	$p^2(1 - p)^2$
SSFS	3	$p^3(1 - p)$	*FSFS*	2	$p^2(1 - p)^2$
SSFF	2	$p^2(1 - p)^2$	*FSFF*	1	$p(1 - p)^3$
SFSS	3	$p^3(1 - p)$	*FFSS*	2	$p^2(1 - p)^2$
SFSF	2	$p^2(1 - p)^2$	*FFSF*	1	$p(1 - p)^3$
SFFS	2	$p^2(1 - p)^2$	*FFFS*	1	$p(1 - p)^3$
SFFF	1	$p(1 - p)^3$	*FFFF*	0	$(1 - p)^4$

En este caso especial, se desea $b(x; 4, p)$ con $x = 0, 1, 2, 3$ y 4. Para $b(3; 4, p)$, identifíquese cuáles de los 16 resultados dan un valor x de 3 y sume las probabilidades asociadas con cada resultado:

$$b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1 - p)$$

Existen cuatro resultados con $X = 3$ y la probabilidad de cada uno es $p^3(1 - p)$(el orden de los S y las F no es importante, sino sólo el número de los S), por lo tanto

$$b(3; 4, p) = \begin{Bmatrix} \text{número de resultados} \\ \text{con } X = 3 \end{Bmatrix} \cdot \begin{Bmatrix} \text{probabilidad de cualquier} \\ \text{resultado con } X = 3 \end{Bmatrix}$$

Asimismo, $b(2; 4, p) = 6p^2(1 - p)^2$, la cual también es el producto del número de resultados con $X = 2$ y la probabilidad de cualquier resultado como ése.

En general,

$$b(x; n, p) = \begin{Bmatrix} \text{número de secuencias de longitud } n \\ \text{compuestas de } x \text{ éxitos} \end{Bmatrix} \cdot \begin{Bmatrix} \text{probabilidad de cualquier} \\ \text{secuencia como ésa} \end{Bmatrix}$$

Como el orden de los S y las F no es importante, el segundo factor en la ecuación previa es $p^x(1 - p)^{n-x}$ (p. ej., los primeros x ensayos producen S y los últimos $n - x$ producen F). El primer factor es el número de formas de escoger x de los n ensayos para que sean los S, es decir, el número de combinaciones de tamaño x que pueden ser construidas con n objetos distintos (ensayos en este caso).

TEOREMA

$$b(x; n, p) = \begin{cases} \dbinom{n}{x} p^x(1 - p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{de lo contrario} \end{cases}$$

Ejemplo 3.31 A cada uno de seis bebedores de refrescos de cola seleccionados al azar se le sirve un vaso de refresco de cola S y uno de refresco de cola F. Los vasos son idénticos en apariencia excepto por un código que viene en el fondo para identificar el refresco de cola. Suponga que en realidad no existe una tendencia entre los bebedores de refresco de cola de preferir un refresco de cola en vez del otro. Entonces $p = P(\text{un individuo seleccionado prefiere } S) = .5$, así que con $X =$ el número entre los seis que prefieren S, $X \sim \text{Bin}(6, .5)$.

Por lo tanto

$$P(X = 3) = b(3; 6, .5) = \binom{6}{3}(.5)^3(.5)^3 = 20(.5)^6 = .313$$

La probabilidad de que por lo menos tres prefieran S es

$$P(3 \le X) = \sum_{x=3}^{6} b(x; 6, .5) = \sum_{x=3}^{6} \binom{6}{x}(.5)^x(.5)^{6-x} = .656$$

y la probabilidad de que cuando mucho uno prefiera S es

$$P(X \le 1) = \sum_{x=0}^{1} b(x; 6, .5) = .109$$

■

Utilización de tablas binomiales*

Incluso con un valor relativamente pequeño de n, el cálculo de probabilidades binomiales es tedioso. La tabla A.1 del apéndice tabula la función de distribución acumulativa $F(x) = P(X \le x)$ con $n = 5, 10, 15, 20, 25$ en combinación con valores seleccionados de p. Varias otras probabilidades pueden entonces ser calculadas por medio de la proposición sobre funciones de distribución acumulativa de la sección 3.2. Una anotación de 0 en la tabla significa únicamente que la probabilidad es 0 a tres dígitos significativos puesto que todos los valores ingresados en la tabla en realidad son positivos.

NOTACIÓN

Para $X \sim \text{Bin}(n, p)$, la función de distribución acumulativa será denotada por

$$B(x; n, p) = P(X \le x) = \sum_{y=0}^{x} b(y; n, p) \qquad x = 0, 1, \ldots, n$$

Ejemplo 3.32 Suponga que 20% de todos los ejemplares de un libro de texto particular no pasan una prueba de resistencia de encuadernación. Sea X el número entre 15 ejemplares seleccionados al azar que no pasan la prueba. Entonces X tiene una distribución binomial con $n = 15$ y $p = .2$.

1. La probabilidad de que cuando mucho 8 no pasen la prueba es

$$P(X \le 8) = \sum_{y=0}^{8} b(y; 15, .2) = B(8; 15, .2)$$

la cual es el dato en el renglón $x = 8$ y la columna $p = .2$ de la tabla binomial $n = 15$. Según la tabla A.1 del apéndice, la probabilidad es $B(8; 15, .2) = .999$.

2. La probabilidad de que exactamente 8 fallen es

$$P(X = 8) = P(X \le 8) - P(X \le 7) = B(8; 15, .2) - B(7; 15, .2)$$

la cual es la diferencia entre dos datos consecutivos en la columna $p = .2$. El resultado es $.999 - .996 = .003$.

* Los paquetes de programas estadísticos tales como Minitab y R proporcionan la función de masa de probabilidad o la función de distribución acumulativa en forma casi instantánea al solicitarla para cualquier valor de p y n desde 2 hasta millones. También existe un comando R para calcular la probabilidad de que X quede en algún intervalo.

3. La probabilidad de que por lo menos 8 fallen es

$$\begin{aligned} P(X \geq 8) &= 1 - P(X \leq 7) = 1 - B(7; 15, .2) \\ &= 1 - \begin{pmatrix} \text{dato en } x = 7 \\ \text{renglón de la columna } p = .2 \end{pmatrix} \\ &= 1 - .996 = .004 \end{aligned}$$

4. Finalmente, la probabilidad de que entre 4 y 7, inclusive, fallen es

$$\begin{aligned} P(4 \leq X \leq 7) &= P(X = 4, 5, 6 \text{ o } 7) = P(X \leq 7) - P(X \leq 3) \\ &= B(7; 15, .2) - B(3; 15, .2) = .996 - .648 = .348 \end{aligned}$$

Obsérvese que esta última probabilidad es la diferencia entre los datos en los renglones $x = 7$ y $x = 3$, *no* en los renglones $x = 7$ y $x = 4$. ■

Ejemplo 3.33 Un fabricante de aparatos electrónicos afirma que cuando mucho 10% de sus unidades de suministro de potencia necesitan servicio durante el periodo de garantía. Para investigar esta afirmación, técnicos en un laboratorio de prueba adquieren 20 unidades y someten a cada una a una prueba acelerada para simular el uso durante el periodo de garantía. Sea p la probabilidad de que una unidad de suministro de potencia necesite reparación durante el periodo (proporción de unidades que requieren reparación). Los técnicos de laboratorio deben decidir si los datos obtenidos con el experimento respaldan la afirmación de que $p \leq .10$. Sea X el número entre las 20 muestreadas que necesitan reparación, así que $X \sim \text{Bin}(20, p)$. Considere la regla de decisión:

Rechazar la afirmación de que $p \leq .10$ a favor de la conclusión de que $p > .10$ si $x \geq 5$ (donde x es el valor observado de X) y considere plausible la afirmación si $x \leq 4$.

La probabilidad de que la afirmación sea rechazada cuando $p = .10$ (una conclusión incorrecta) es

$$P(X \geq 5 \text{ cuando } p = .10) = 1 - B(4; 20, .1) = 1 - .957 = .043$$

La probabilidad de que la afirmación no sea rechazada cuando $p = .20$ (un tipo diferente de conclusión incorrecta) es

$$P(X \leq 4 \text{ cuando } p = .2) = B(4; 20, .2) = .630$$

La primera probabilidad es algo pequeña, pero la segunda es intolerablemente grande. Cuando $p = .20$, significa que el fabricante subestimó de manera excesiva el porcentaje de unidades que necesitan servicio, y si se utiliza la regla de decisión establecida, ¡el 63% de todas las muestras resultaron plausibles!

Se podría pensar que la probabilidad de este segundo tipo de conclusión errónea podría hacerse más pequeña cambiando el valor de corte de 5 en la regla de decisión a algo más. Sin embargo, aunque el reemplazo de 5 por un número más pequeño daría una probabilidad más pequeña que .630, la otra probabilidad se incrementaría entonces. La única forma de hacer ambas “probabilidades de error” pequeñas es basar la regla de decisión en un experimento que implique muchas más unidades. ■

La media y varianza de X

Con $n = 1$, la distribución binomial llega a ser la distribución de Bernoulli. De acuerdo con el ejemplo 3.18, el valor medio de una variable de Bernoulli es $\mu = p$, así que el número esperado de las S en cualquier ensayo único es p. Como un experimento binomial se compone de n ensayos, la intuición sugiere que para $X \sim \text{Bin}(n, p)$, $E(X) = np$, el producto del número de ensayos y la probabilidad de éxito en un solo ensayo. La expresión para $V(X)$ no es tan intuitiva.

PROPOSICIÓN

Si $X \sim \text{Bin}(n, p)$, entonces $E(X) = np$, $V(X) = np(1 - p) = npq$ y $\sigma_X = \sqrt{npq}$ (donde $q = 1 - p$).

Por tanto, para calcular la media y varianza de una variable aleatoria binomial no se requiere evaluar las sumas. La comprobación del resultado para $E(X)$ se ilustra en el ejercicio 64.

Ejemplo 3.34 Si 75% de todas las compras en una tienda se hacen con tarjeta de crédito y X es el número entre diez compras seleccionadas al azar realizadas con tarjeta de crédito, entonces $X \sim \text{Bin}(10, .75)$. Por lo tanto, $E(X) = np = (10)(.75) = 7.5$, $V(X) = npq = 10(.75)(.25) = 1.875$ y $\sigma = \sqrt{1.875} = 1.37$. Otra vez, aun cuando X puede tomar sólo valores enteros, $E(X)$ no tiene que ser un entero. Si se realiza un gran número de experimentos binomiales independientes, cada uno con $n = 10$ ensayos y $p = .75$, entonces el número promedio de las S por experimento se acercará a 7.5.

La probabilidad de que X se encuentre dentro de una desviación estándar de su valor medio es $P(7.5 - 1.37 \le X \le 7.5 + 1.37) = P(6.13 \le X \le 8.87) = P(X = 7 \text{ u } 8) = .532$. ■

EJERCICIOS Sección 3.4 (46–67)

46. Calcule las siguientes probabilidades binomiales directamente con la fórmula para $b(x; n, p)$:
a. $b(3; 8, .35)$
b. $b(5; 8, .6)$
c. $P(3 \le X \le 5)$ cuando $n = 7$ y $p = .6$
d. $P(1 \le X)$ cuando $n = 9$ y $p = .1$

47. Use la tabla A.1 del apéndice para obtener las siguientes probabilidades:
a. $B(4; 15, .3)$
b. $b(4; 15, .3)$
c. $b(6; 15, .7)$
d. $P(2 \le X \le 4)$ cuando $X \sim \text{Bin}(15, .3)$
e. $P(2 \le X)$ cuando $X \sim \text{Bin}(15, .3)$
f. $P(X \le 1)$ cuando $X \sim \text{Bin}(15, .7)$
g. $P(2 < X < 6)$ cuando $X \sim \text{Bin}(15, .3)$

48. Cuando las tarjetas de circuito usadas en la fabricación de reproductores de discos compactos se prueban, el porcentaje de defectuosas es de 5%. Sea $X =$ el número de tarjetas defectuosas en una muestra aleatoria de tamaño $n = 25$, así que $X \sim \text{Bin}(25, .05)$.
a. Determine $P(X \le 2)$.
b. Determine $P(X \ge 5)$.
c. Determine $P(1 \le X \le 4)$.
d. ¿Cuál es la probabilidad de que ninguna de estas 25 tarjetas esté defectuosa?
e. Calcule el valor esperado y la desviación estándar de X.

49. Una compañía que produce cristal fino sabe por experiencia que 10% de sus copas de mesa tienen imperfecciones cosméticas y deben ser clasificadas como "segundas".
a. Entre seis copas seleccionadas al azar, ¿qué tan probable es que sólo una sea segunda?
b. Entre seis copas seleccionadas al azar, ¿qué tan probable es que por lo menos dos sean segundas?
c. Si las copas se examinan una por una, ¿cuál es la probabilidad de que cuando mucho cinco deban ser seleccionadas para encontrar cuatro que no sean segundas?

50. Se utiliza un número telefónico particular para recibir tanto llamadas de voz como faxes. Suponga que 25% de las llamadas entrantes son faxes y considere una muestra de 25 llamadas entrantes. ¿Cuál es la probabilidad de que
a. cuando mucho 6 de las llamadas sean un fax?
b. exactamente 6 de las llamadas sean un fax?
c. por lo menos 6 de las llamadas sean un fax?
d. más de 6 de las llamadas sean un fax?

51. Remítase al ejercicio previo.
a. ¿Cuál es el número esperado de llamadas entre las 25 que implican un fax?
b. ¿Cuál es la desviación estándar del número entre las 25 llamadas que implican un fax?
c. ¿Cuál es la probabilidad de que el número de llamadas entre las 25 que implican una transmisión de fax sobrepase el número esperado por más de 2 desviaciones estándar?

52. Suponga que 30% de todos los estudiantes que tienen que comprar un texto para un curso particular desean un ejemplar nuevo (¡los exitosos!), mientras que el otro 70% desea comprar un ejemplar usado. Considere seleccionar 25 compradores al azar.
a. ¿Cuáles son el valor medio y la desviación estándar del número que desea un ejemplar nuevo del libro?
b. ¿Cuál es la probabilidad de que el número que desea ejemplares nuevos esté a más de dos desviaciones estándar del valor medio?

c. La librería tiene 15 ejemplares nuevos y 15 usados en existencia. Si 25 personas llegan una por una a comprar el texto, ¿cuál es la probabilidad de que las 25 obtengan el tipo de libro que desean de las existencias actuales? [*Sugerencia*: sea $X =$ el número que desea un ejemplar nuevo. ¿Con qué valores de X obtendrán las 25 personas lo que desean?]

d. Suponga que los ejemplares nuevos cuestan $100 y los usados $70. Suponga que la librería en la actualidad tiene 50 ejemplares nuevos y 50 usados. ¿Cuál es el valor esperado del ingreso total por la venta de los siguientes 25 ejemplares comprados? Asegúrese de indicar qué regla de valor esperado está utilizando. [*Sugerencia*: sea $h(X) =$ el ingreso cuando X de los 25 compradores desean ejemplares nuevos. Exprese esto como una función lineal.]

53. El ejercicio 30 (sección 3.3) dio la función de masa de probabilidad de Y, el número de infracciones de tránsito de un individuo seleccionado al azar asegurado por una compañía particular. ¿Cuál es la probabilidad de que entre 15 individuos seleccionados al azar

a. por lo menos 10 no tengan infracciones?

b. menos de la mitad tengan por lo menos una infracción?

c. el número que tengan por lo menos una infracción esté entre 5 y 10, inclusive?*

54. Un tipo particular de raqueta de tenis viene en tamaño mediano y en tamaño extragrande. Sesenta por ciento de todos los clientes en una tienda desean la versión extragrande.

a. Entre diez clientes seleccionados al azar que desean este tipo de raqueta, ¿cuál es la probabilidad de que por lo menos seis deseen la versión extragrande?

b. Entre diez clientes seleccionados al azar, ¿cuál es la probabilidad de que el número que desea la versión extragrande esté dentro de 1 desviación estándar del valor medio?

c. La tienda dispone actualmente de siete raquetas de cada versión. ¿Cuál es la probabilidad de que los siguientes diez clientes que desean esta raqueta puedan obtener la versión que desean de las existencias actuales?

55. Veinte por ciento de todos los teléfonos de cierto tipo son llevados a servicio mientras se encuentran dentro de la garantía. De éstos, 60% pueden ser reparados, mientras el 40% restante deben ser reemplazados con unidades nuevas. Si una compañía adquiere diez de estos teléfonos, ¿cuál es la probabilidad de que exactamente dos sean reemplazados en la vigencia de su garantía?

56. La Junta de Educación reporta que 2% de los dos millones de estudiantes de preparatoria que toman el examen de aptitud escolar cada año reciben un trato especial a causa de discapacidades documentadas (*Los Ángeles Times*, 16 de julio de 2002). Considere una muestra aleatoria de 25 estudiantes que recientemente presentaron el examen.

a. ¿Cuál es la probabilidad de que exactamente 1 reciba un trato especial?

b. ¿Cuál es la probabilidad de que por lo menos 1 reciba un trato especial?

c. ¿Cuál es la probabilidad de que por lo menos 2 reciban un trato especial?

d. ¿Cuál es la probabilidad de que el número entre los 25 que recibieron un trato especial esté dentro de 2 desviaciones estándar del número que esperaría reciba un trato especial?

e. Suponga que a un estudiante que no recibe un trato especial se le permiten 3 horas para el examen, mientras que a un estudiante que recibió un trato especial se le permiten 4.5 horas. ¿Qué tiempo promedio piensa que le sería permitido a los 25 estudiantes seleccionados?

57. Suponga que 90% de todas las baterías de cierto proveedor tienen voltajes aceptables. Un tipo de linterna requiere que las dos baterías sean tipo D y funcionará sólo si sus dos baterías tienen voltajes aceptables. Entre diez linternas seleccionadas al azar, ¿cuál es la probabilidad de que por lo menos nueve funcionarán? ¿Qué suposiciones hizo para responder la pregunta planteada?

58. Un distribuidor recibe un lote muy grande de componentes. El lote sólo puede ser caracterizado como aceptable si la proporción de componentes defectuosos es cuando mucho de .10. El distribuidor decide seleccionar 10 componentes al azar y aceptar el lote sólo si el número de componentes defectuosos presentes en la muestra es cuando mucho de 2.

a. ¿Cuál es la probabilidad de que el lote será aceptado cuando la proporción real de componentes defectuosos es de .01?, ¿.05? ¿.10? ¿.20? ¿.25?

b. Sea p la proporción real de componentes defectuosos presentes en el lote. Una gráfica de P(se acepta el lote) en función de p, con p sobre el eje horizontal y P(se acepta el lote) sobre el eje vertical, se llama *curva característica de operación* del plan de muestreo de aceptación. Use los resultados del inciso (a) para trazar esta curva con $0 \leq p \leq 1$.

c. Repita los incisos (a) y (b) con "1" reemplazando a "2" en el plan de muestreo de aceptación.

d. Repita los incisos (a) y (b) con "15" reemplazando a "10" en el plan de muestreo de aceptación.

e. ¿Cuál de estos planes de muestreo, el del inciso (a), (c) o (d) parece más satisfactorio y por qué?

59. Un reglamento que requiere que se instale un detector de humo en todas las casas previamente construidas ha estado en vigor en una ciudad particular durante 1 año. Al departamento de bomberos le preocupa que muchas casas permanezcan sin detectores. Sea $p =$ la proporción verdadera de las casas que tienen detectores y suponga que se inspecciona una muestra aleatoria de 25 casas. Si ésta indica marcadamente que menos de 80% de todas las casas tienen un detector, el departamento de bomberos lanzará una campaña para la puesta en ejecución de un programa de inspección obligatorio. Debido a lo caro del programa, el departamento prefiere no requerir tales inspecciones a menos que una evidencia muestral indique que se requieren. Sea X el número de casas con detectores entre las 25 muestreadas. Considere rechazar el requerimiento de que $p \geq .8$ si $x \leq 15$.

a. ¿Cuál es la probabilidad de que el requerimiento sea rechazado cuando el valor real de p es .8?

b. ¿Cuál es la probabilidad de no rechazar el requerimiento cuando $p = .7$? ¿Cuando $p = .6$?

c. ¿Cómo cambian las "probabilidades de error" de los incisos (a) y (b) si el valor 15 en la regla de decisión es reemplazado por 14?

* "Entre a y b, inclusive" equivale a ($a \leq X \leq b$).

60. Un puente de peaje cobra \$ 1.00 para los vehículos de pasajeros y \$ 2.50 para los demás vehículos. Supongamos que durante las horas del día, el 60% de todos los vehículos son de pasajeros. Si 25 vehículos cruzan el puente durante un periodo determinado durante el día, ¿cuál es el resultado de los ingresos por peaje previstos? [*Sugerencia*: Sea $X =$ el número de vehículos de pasajeros, entonces, los ingresos por peaje $h(X)$ son una función lineal de X.]

61. Un estudiante que está tratando de escribir un ensayo para un curso tiene la opción de dos temas, A y B. Si selecciona el tema A, el estudiante pedirá dos libros mediante préstamo interbiblioteca, mientras que si selecciona el tema B, el estudiante pedirá cuatro libros. El estudiante cree que un buen ensayo necesita recibir y utilizar por lo menos la mitad de los libros pedidos para uno u otro tema seleccionado. Si la probabilidad de que un libro pedido mediante préstamo interbiblioteca llegue a tiempo es de .9 y los libros llegan independientemente uno de otro, ¿qué tema deberá seleccionar el estudiante para incrementar al máximo la probabilidad de escribir un buen ensayo? ¿Qué pasa si la probabilidad de que lleguen los libros es de sólo .5 en lugar de .9?

62. **a.** Con n fijo, ¿hay valores de $p\,(0 \le p \le 1)$ para los cuales $V(X) = 0$? Explique por qué esto es así.

b. ¿Con qué valor de p se incrementa al máximo $V(X)$? [*Sugerencia*: grafique $V(X)$ en función de p o bien saque una derivada.]

63. **a.** Demuestre que $b(x; n, 1 - p) = b(n - x; n, p)$.

b. Demuestre que $B(x; n, 1 - p) = 1 - B(n - x - 1; n, p)$. [*Sugerencia*: cuando mucho el número x de los S equivale a por lo menos $(n - x)$ de las F.]

c. ¿Qué implican los incisos (a) y (b) sobre la necesidad de incluir valores de p más grandes que .5 en la tabla A.1 del apéndice?

64. Demuestre que $E(X) = np$ cuando X es una variable aleatoria binomial. [*Sugerencia*: primero exprese $E(X)$ como una suma con límite inferior $x = 1$. Luego saque a np como factor, sea $y = x - 1$ de modo que la suma sea de $y = 0$ a $y = n - 1$ y demuestre que la suma es igual a 1.]

65. Los clientes en una gasolinera pagan con tarjeta de crédito (A), tarjeta de débito (B) o efectivo (C). Suponga que clientes sucesivos toman decisiones independientes con $P(A) = .5$, $P(B) = .2$ y $P(C) = .3$.

a. Entre los siguientes 100 clientes, ¿cuáles son la media y la varianza del número que paga con tarjeta de débito? Explique su razonamiento.

b. Conteste el inciso (a) para el número entre los 100 que no pagan con efectivo.

66. Una limusina de aeropuerto puede transportar hasta cuatro pasajeros en cualquier viaje. La compañía aceptará un máximo de seis reservaciones para un viaje y un pasajero debe tener una reservación. Según registros previos, 20% de los que reservan no se presentan para el viaje. Responda las siguientes preguntas, suponiendo independencia en los casos en que sea apropiado.

a. Si se hacen seis reservaciones, ¿cuál es la probabilidad de que por lo menos un individuo con reservación no pueda ser acomodado en el viaje?

b. Si se hacen seis reservaciones, ¿cuál es el número esperado de lugares disponibles cuando la limusina parte?

c. Suponga que la distribución de probabilidad del número de reservaciones hechas se da en la tabla adjunta.

Número de reservaciones	3	4	5	6
Probabilidad	.1	.2	.3	.4

Sea X el número de pasajeros en un viaje seleccionado al azar. Obtenga la función de masa de probabilidad de X.

67. Remítase a la desigualdad de Chebyshev dada en el ejercicio 44. Calcule $P(|X - \mu| \ge k\sigma)$ con $k = 2$ y $k = 3$ cuando $X \sim \text{Bin}(20, .5)$ y compare con el límite superior correspondiente. Repita para $X \sim \text{Bin}(20, .75)$.

3.5 Distribuciones hipergeométrica y binomial negativa

Las distribuciones hipergeométrica y binomial negativa están relacionadas con la distribución binomial. La distribución binomial es el modelo de probabilidad aproximada de muestreo sin reemplazo de una población dicotómica finita (S–F). Si el tamaño n de la muestra es pequeño con respecto al tamaño N de la población, la distribución hipergeométrica es el modelo de probabilidad exacta del número de éxitos (S) en la muestra. La variable aleatoria binomial X es el número de los S cuando el número n de ensayos es fijo, mientras la distribución binomial surge de fijar el número de éxitos deseados y de permitir que el número de ensayos sea aleatorio.

Distribución hipergeométrica

Las suposiciones que conducen a la distribución hipergeométrica son las siguientes:

1. La población o conjunto que se va a muestrear se compone de N individuos, objetos o elementos (una población *finita*).

2. Cada individuo puede ser caracterizado como éxito (S) o falla (F) y hay M éxitos en la población.
3. Se selecciona una muestra de n individuos sin reemplazo de tal modo que cada subconjunto de tamaño n tenga la misma probabilidad de ser seleccionado.

La variable aleatoria de interés es $X =$ el número de las S en la muestra. La distribución de probabilidad de X depende de los parámetros n, M y N, así que se desea obtener $P(X = x) = h(x; n, M, N)$.

Ejemplo 3.35 Durante un periodo particular una oficina de tecnología de la información de una universidad recibió 20 solicitudes de servicio por problemas con impresoras, de las cuales 8 eran impresoras láser y 12 eran modelos de inyección de tinta. Se tiene que seleccionar una muestra de 5 de estas solicitudes de servicio para incluirla en una encuesta sobre satisfacción del cliente. Suponga que las 5 son seleccionadas completamente al azar, de modo que cualquier subconjunto de tamaño 5 tenga la misma probabilidad de ser seleccionado como cualquier otro subconjunto. ¿Cuál es entonces la probabilidad de que exactamente $x(x = 0, 1, 2, 3, 4$ o $5)$ de las solicitudes de servicio seleccionadas sean para impresoras de inyección de tinta?

En este caso, el tamaño de la población es $N = 20$, el tamaño de la muestra es $n = 5$ y el número de éxitos (inyección de tinta $= S$) y las fallas (F) en la población son $M = 12$ y $N - M = 8$, respectivamente. Considérese el valor $x = 2$. Como todos los resultados (cada uno de los cuales consta de 5 solicitudes particulares) son igualmente probables,

$$P(X = 2) = h(2; 5, 12, 20) = \frac{\text{número de resultados con } X = 2}{\text{número de resultados posibles}}$$

El número de resultados posibles en el experimento es el número de formas de seleccionar 5 de los 20 objetos sin importar el orden, es decir, $\binom{20}{5}$. Para contar el número de resultados con $X = 2$, obsérvese que existen $\binom{12}{2}$ formas de seleccionar 2 de las solicitudes para impresoras de inyección de tinta, y por cada forma existen $\binom{8}{3}$ formas de seleccionar las 3 solicitudes para impresoras láser a fin de completar la muestra. La regla de producto del capítulo 2 da entonces $\binom{12}{2}\binom{8}{3}$ como el número de resultados con $X = 2$, por lo tanto

$$h(2; 5, 12, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = .238 \qquad \blacksquare$$

En general, si el tamaño de la muestra n es más pequeño que el número de éxitos en la población (M), entonces el valor de X más grande posible es n. Sin embargo, si $M < n$ (p. ej., un tamaño de muestra de 25 y sólo hay 15 éxitos en la población), entonces X puede ser cuando mucho M. Asimismo, siempre que el número de fallas en la población ($N - M$) sobrepase el tamaño de la muestra, el valor más pequeño posible de X es 0 (puesto que todos los individuos muestreados podrían entonces ser fallas). Sin embargo, si $N - M < n$, el valor más pequeño posible de X es $n - (N - M)$. Por lo tanto, los posibles valores de X satisfacen la restricción máx $(0, n - (N - M)) \le x \le$ mín (n, M). Un argumento paralelo al del ejemplo previo da la función de masa de probabilidad de X.

PROPOSICIÓN

Si X es el número de éxitos (S) en una muestra completamente aleatoria de tamaño n extraída de la población compuesta de M éxitos y $(N - M)$ fallas, entonces la distribución de probabilidad de X, llamada **distribución hipergeométrica**, es

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (3.15)$$

con x, un entero, que satisface máx $(0, n - N + M) \le x \le$ mín (n, M).

En el ejemplo 3.35, $n = 5$, $M = 12$ y $N = 20$, por lo tanto $h(x; 5, 12, 20)$ con $x = 0, 1, 2, 3, 4, 5$ se obtiene sustituyendo estos números en la ecuación (3.15).

Ejemplo 3.36 Se capturaron, etiquetaron y liberaron cinco individuos de una población de animales que se piensa están al borde la extinción en una región para que se mezclen con la población. Después de haber tenido la oportunidad de mezclarse, se selecciona una muestra aleatoria de 10 de estos animales. Sea X = el número de animales etiquetados en la segunda muestra. Si en realidad hay 25 animales de este tipo en la región, ¿cuál es la probabilidad de que (a) $X = 2$? (b) $X \leq 2$?

Los valores de los parámetros son $n = 10$, $M = 5$ (5 animales etiquetados en la población) y $N = 25$, por lo tanto

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}} \qquad x = 0, 1, 2, 3, 4, 5$$

Para el inciso (a),

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

Para el inciso (b),

$$\begin{aligned} P(X \leq 2) = P(X = 0, 1 \text{ o } 2) &= \sum_{x=0}^{2} h(x; 10, 5, 25) \\ &= .057 + .257 + .385 = .699 \end{aligned}$$

■

Varios paquetes de software estadístico generan fácilmente probabilidades hipergeométricas (tabular es engorroso, a causa de los tres parámetros).

Como en el caso binomial, existen expresiones simples para $E(X)$ y $V(X)$ para variables aleatorias hipergeométricas.

PROPOSICIÓN

La media y la varianza de la variable aleatoria hipergeométrica X cuya función de masa de probabilidad es $h(x; n, M, N)$ son

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

La razón M/N es la proporción de éxitos en la población. Si se reemplaza M/N por p en $E(X)$ y $V(X)$, se obtiene

$$\begin{aligned} E(X) &= np \\ V(X) &= \left(\frac{N-n}{N-1}\right) \cdot np(1-p) \end{aligned} \qquad (3.16)$$

La expresión (3.16) muestra que las medias de las variables aleatorias binomiales e hipergeométricas son iguales, en tanto que las varianzas de las dos variables aleatorias difieren por el factor $(N - n)/(N - 1)$, a menudo llamado **factor de corrección de población finita.** Este factor es menor que 1, así que la variable hipergeométrica tiene una varianza más pequeña que la variable aleatoria binomial. El factor de corrección puede escribirse como $(1 - n/N)/(1 - 1/N)$, el cual es aproximadamente 1 cuando n es pequeño con respecto a N.

Ejemplo 3.37 (Continuación del ejemplo 3.36)

En el ejemplo de etiquetación de animales, $n = 10$, $M = 5$ y $N = 25$, por lo tanto $p = \frac{5}{25} = .2$ y

$$E(X) = 10(.2) = 2$$

$$V(X) = \frac{15}{24}(10)(.2)(.8) = (.625)(1.6) = 1$$

Si el muestreo se realizó con reemplazo, $V(X) = 1.6$.

Suponga que en realidad no se conoce el tamaño de la población N, así que se observa el valor x y se desea estimar N. Es razonable igualar la proporción muestral observada de éxitos, x/n, con la proporción de la población, M/N, que da la estimación

$$\hat{N} = \frac{M \cdot n}{x}$$

Si $M = 100$, $n = 40$ y $x = 16$, entonces $\hat{N} = 250$. ■

La regla general empírica dada en la sección 3.4 plantea que si el muestreo se realizó sin reemplazo pero n/N era cuando mucho de .05, entonces la distribución binomial podría ser utilizada para calcular probabilidades aproximadas que implican el número de éxitos en la muestra. Un enunciado más preciso es el siguiente: permita que el tamaño de la población N y el número de M éxitos presentes en la población se hagan más grandes a medida que la razón M/N tiende a p. Entonces $h(x; n, M, N)$ tiende a $b(x; n, p)$; así que con n/N pequeña, las dos son aproximadamente iguales siempre y cuando p no esté muy cerca de 0 o 1. Éste es la razón de ser de la regla empírica.

Distribución binomial negativa

La variable aleatoria binomial y la distribución binomial negativa se basan en un experimento que satisface las siguientes condiciones:

1. El experimento consiste en una secuencia de ensayos independientes.
2. Cada ensayo puede dar por resultado un éxito (S) o una falla (F).
3. La probabilidad de éxito es constante de un ensayo a otro, por lo tanto $P(S$ en el ensayo $i) = p$ con $i = 1, 2, 3, \ldots$
4. El experimento continúa (se realizan ensayos) hasta que un total de r éxitos hayan sido observados, donde r es un entero positivo especificado.

La variable aleatoria de interés es $X =$ el número de fallas que preceden al éxito r-ésimo; X se llama **variable aleatoria binomial negativa** porque, en contraste con la variable aleatoria binomial, el número de éxitos es fijo y el número de ensayos es aleatorio.

Posibles valores de X son $0, 1, 2, \ldots$ Sea $nb(x; r, p)$ la función de masa de probabilidad de X. Considere $nb(7, 3, p) = P(X = 7)$, la probabilidad de que ocurran exactamente 7F antes de la 3ª S. Para que esto suceda, el décimo ensayo debe ser una S y debe haber exactamente 2 S entre los 9 primeros ensayos. Por tanto

$$nb(7; 3, p) = \left\{\binom{9}{2} \cdot p^2(1 - p)^7\right\} \cdot p = \binom{9}{2} \cdot p^3(1 - p)^7$$

La generalización de esta línea de razonamiento da la siguiente fórmula para la función de masa de probabilidad binomial negativa.

PROPOSICIÓN

La función de masa de probabilidad de la variable aleatoria binomial negativa X con los parámetros r = número de éxitos (S) y $p = P(S)$ es

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \ldots$$

Ejemplo 3.38 Un pediatra desea reclutar 5 parejas, cada una de las cuales espera a su primer hijo, para participar en un nuevo régimen de alumbramiento natural. Sea $p = P$(una pareja seleccionada al azar está de acuerdo en participar). Si $p = .2$, ¿cuál es la probabilidad de que 15 parejas tengan que ser entrevistadas antes de encontrar 5 que estén de acuerdo en participar? Es decir, con $S = \{$está de acuerdo en participar$\}$, ¿cuál es la probabilidad de que ocurran 10 fallas antes del quinto éxito? Sustituyendo $r = 5$, $p = .2$ y $x = 10$ en $nb(x; r, p)$ da

$$nb(10; 5, .2) = \binom{14}{4} (.2)^5 (.8)^{10} = .034$$

La probabilidad de que cuando mucho se observen 10 fallas (cuando mucho con 15 parejas entrevistadas) es

$$P(X \leq 10) = \sum_{x=0}^{10} nb(x; 5, .2) = (.2)^5 \sum_{x=0}^{10} \binom{x + 4}{4} (.8)^x = .164$$ ■

En algunas fuentes, la variable aleatoria binomial negativa se considera como el número de ensayos $X + r$ en lugar del número de fallas.

En el caso especial $r = 1$, la función de masa de probabilidad es

$$nb(x; 1, p) = (1 - p)^x p \quad x = 0, 1, 2, \ldots \tag{3.17}$$

En el ejemplo 3.12 se dedujo la función de masa de probabilidad para el número de ensayos necesarios para obtener el primer *éxito* (S) y allí la función de masa de probabilidad es similar a la expresión (3.17). En la literatura se hace referencia tanto a X = número de *fallas* (F) como a Y = número de ensayos ($= 1 + X$) como **variables aleatorias geométricas,** y la función de masa de probabilidad en la expresión (3.17) se llama **distribución geométrica.**

En el ejemplo 3.19 se demostró que el número esperado de ensayos hasta que aparece el primer éxito es $1/p$, así que el número esperado de *fallas* hasta que aparece el primer éxito es $(1/p) - 1 = (1 - p)/p$. Intuitivamente, se esperaría ver $r \cdot (1 - p)/p$ fallas antes del éxito r-ésimo y éste en realidad es $E(X)$. También existe una fórmula simple para $V(X)$.

PROPOSICIÓN

Si X es una variable aleatoria binomial negativa con función de masa de probabilidad $nb(x; r, p)$, entonces

$$E(X) = \frac{r(1 - p)}{p} \qquad V(X) = \frac{r(1 - p)}{p^2}$$

Por último, al expandir el coeficiente binomial enfrente de $p^r(1 - p)^x$ y haciendo alguna cancelación, se ve que $nb(x; r, p)$ está bien definido incluso cuando r no es un entero. Se ha encontrado que la *distribución binomial negativa generalizada* para ajustar los datos observados verdaderamente bien en una amplia variedad de aplicaciones.

EJERCICIOS Sección 3.5 (68–78)

68. Una tienda de electrónica ha recibido un envío de 20 radios de mesa que tienen conexiones para el iPod o iPhone. Doce de ellos tienen dos ranuras (para que puedan acomodar a los dos dispositivos) y los otros ocho tienen una sola ranura. Supongamos que seis de los 20 radios son seleccionados al azar para ser almacenados en un estante donde son exhibidos y los restantes se colocan en un almacén. Sea X = el número de los radios almacenados en el estante de exhibición que tienen dos ranuras.

- **a.** ¿Qué clase de distribución tiene X (nombre y valores de todos los parámetros)?
- **b.** Calcule $P(X = 2)$, $P(X \leq 2)$ y $P(X \geq 2)$.
- **c.** Calcule el valor medio y la desviación estándar de X.

69. Cada uno de 12 refrigeradores de un tipo ha sido regresado a un distribuidor debido a un ruido agudo audible producido por oscilación cuando el refrigerador está funcionando. Suponga que 7 de estos refrigeradores tienen un compresor defectuoso y que los otros 5 tienen problemas menos serios. Si los refrigeradores se examinan en orden aleatorio, sea X el número entre los primeros 6 examinados que tienen un compresor defectuoso. Calcule lo siguiente:

- **a.** $P(X = 5)$
- **b.** $P(X \leq 4)$
- **c.** La probabilidad de que X exceda su valor medio por más de 1 desviación estándar.
- **d.** Considere un gran envío de 400 refrigeradores, 40 de los cuales tienen compresores defectuosos. Si X es el número entre 15 refrigeradores seleccionados al azar que tienen compresores defectuosos, describa una forma menos tediosa de calcular (por lo menos de forma aproximada) $P(X \leq 5)$ que utilizar la función de masa de probabilidad hipergeométrica.

70. Un instructor que impartió dos secciones de estadística para ingeniería el semestre pasado, la primera con 20 estudiantes y la segunda con 30, decidió asignar un proyecto semestral. Una vez que todos los proyectos le fueron entregados, el instructor los ordenó al azar antes de calificarlos. Considere los primeros 15 proyectos calificados.

- **a.** ¿Cuál es la probabilidad de que exactamente 10 de éstos sean de la segunda sección?
- **b.** ¿Cuál es la probabilidad de que por lo menos 10 de éstos sean de la segunda sección?
- **c.** ¿Cuál es la probabilidad de que por lo menos 10 de éstos sean de la misma sección?
- **d.** ¿Cuáles son el valor medio y la desviación estándar del número entre estos 15 que son de la segunda sección?
- **e.** ¿Cuáles son el valor medio y la desviación estándar del número de proyectos que no están entre estos primeros 15 que son de la segunda sección?

71. Un geólogo recolectó 10 especímenes de roca basáltica y 10 de granito. Él le pide a su ayudante de laboratorio que seleccione al azar 15 de los especímenes para analizarlos.

- **a.** ¿Cuál es la función de masa de probabilidad del número de especímenes de granito seleccionados para su análisis?
- **b.** ¿Cuál es la probabilidad de que todos los especímenes de uno de los dos tipos de roca sean seleccionados para su análisis?
- **c.** ¿Cuál es la probabilidad de que el número de especímenes de granito seleccionados para analizarlos esté dentro de 1 desviación estándar de su valor medio?

72. Un director de personal que va a entrevistar a 11 ingenieros para cuatro vacantes de trabajo ha programado seis entrevistas para el primer día y cinco para el segundo. Suponga que los candidatos son entrevistados en orden aleatorio.

- **a.** ¿Cuál es la probabilidad de que x de los cuatro mejores candidatos sean entrevistados el primer día?
- **b.** ¿Cuántos de los mejores cuatro candidatos se espera que puedan ser entrevistados el primer día?

73. Veinte parejas de individuos que participan en un torneo de bridge han sido sembrados del 1, . . . , 20. En esta primera parte del torneo, los 20 son divididos al azar en 10 parejas este–oeste y 10 parejas norte–sur.

- **a.** ¿Cuál es la probabilidad de que x de las 10 mejores parejas terminen jugando este–oeste?
- **b.** ¿Cuál es la probabilidad de que las cinco mejores parejas terminen jugando en la misma dirección?
- **c.** Si existen $2n$ parejas, ¿cuál es la función de masa de probabilidad de X = el número entre las mejores n parejas que terminan jugando este–oeste? ¿Cuáles son $E(X)$ y $V(X)$?

74. Una alerta contra smog de segunda etapa ha sido emitida en un área del condado de Los Ángeles en la cual hay 50 firmas industriales. Un inspector visitará 10 firmas seleccionadas al azar para ver si no han violado los reglamentos.

- **a.** Si 15 de las firmas sí están violando por lo menos un reglamento, ¿cuál es la función de masa de probabilidad del número de firmas visitadas por el inspector que violan por lo menos un reglamento?
- **b.** Si existen 500 firmas en el área, 150 de las cuales violan algún reglamento, represente de forma aproximada la función de masa de probabilidad del inciso (a) con una función de masa de probabilidad más simple.
- **c.** Con X = el número entre las 10 visitadas que violan algún reglamento, calcule $E(X)$ y $V(X)$ ambas para la función de masa de probabilidad exacta y la función de masa de probabilidad aproximada del inciso (b).

75. Suponga que $p = P(\text{nacimiento de un varón}) = .5$. Una pareja desea tener exactamente dos niñas en su familia. Tendrán hijos hasta que esta condición se satisfaga.

- **a.** ¿Cuál es la probabilidad de que la familia tenga x varones?
- **b.** ¿Cuál es la probabilidad de que la familia tenga cuatro hijos?
- **c.** ¿Cuál es la probabilidad de que la familia tenga cuando mucho cuatro hijos?
- **d.** ¿Cuántos varones cree que tendría esta familia? ¿Cuántos hijos esperaría que tenga esta familia?

76. Una familia decide tener hijos hasta que tengan tres del mismo sexo. Suponiendo $P(B) = P(G) = .5$, ¿cuál es la función de masa de probabilidad de X = el número de hijos en la familia?

77. Tres hermanos y sus esposas deciden tener hijos hasta que cada familia tenga dos niñas. ¿Cuál es la función de masa de probabilidad de X = el número total de varones procreados por los hermanos? ¿Cuál es $E(X)$ y cómo se compara con el número esperado de varones procreados por cada hermano?

78. De acuerdo con el artículo "Characterizing the Severity and Risk of Drought in the Poudre River, Colorado" (*J. of Water Res. Planning and Mgmnt.*, 2005: 383–393), la longitud de la sequía Y es el número de intervalos de tiempo consecutivos en los que el suministro de agua se mantiene por debajo de un valor crítico y_0 (un déficit), precedido y seguido por periodos en los que el suministro supera este valor crítico (un excedente). El documento citado propone una distribución geométrica con $p = .409$ para esta variable aleatoria.

a. ¿Cuál es la probabilidad de que una sequía dure exactamente 3 intervalos? ¿A lo más 3 intervalos?

b. ¿Cuál es la probabilidad de que la duración de una sequía exceda su valor medio por al menos una desviación estándar?

3.6 Distribución de probabilidad de Poisson

Las distribuciones binomial, hipergeométrica y binomial negativa se dedujeron partiendo de un experimento compuesto de ensayos o sorteos y aplicando las leyes de probabilidad a varios resultados del experimento. No existe un experimento simple en el cual esté basada la distribución de Poisson, aun cuando en breve se describirá cómo puede ser obtenida mediante ciertas operaciones restrictivas.

DEFINICIÓN

Se dice que una variable aleatoria discreta X tiene una **distribución de Poisson** con parámetro μ ($\mu > 0$) si la función de masa de probabilidad de X es

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \qquad x = 0, 1, 2, 3, \ldots$$

No es casualidad que se esté usando el símbolo μ para el parámetro de Poisson, en breve se verá que μ es en realidad el valor esperado de X. La letra e en la función de masa de probabilidad representa la base del sistema de logaritmos naturales, su valor numérico es aproximadamente 2.71828. A diferencia de las distribuciones binomial e hipergeométrica, la distribución de probabilidad de Poisson se extiende a *todos* los números enteros no negativos, un número infinito de posibilidades.

No es evidente por inspección que $p(x; \mu)$ especifique una función de masa de probabilidad legítima, por no hablar de que esta distribución es útil. En primer lugar, $p(x; \mu) > 0$ para cada valor x posible debido a la exigencia de que $\mu > 0$. El hecho de que $\sum p(x; \mu) = 1$ es una consecuencia de la expansión en series de Maclaurin de e^{μ} (consulte su libro de cálculo para este resultado):

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \tag{3.18}$$

Si los dos términos extremos de la expresión (3.18) se multiplican por $e^{-\mu}$ y luego esta cantidad se coloca adentro de la suma en el lado derecho, el resultado es

$$1 = \sum_{x=0}^{\infty} \frac{e^{-\mu} \cdot \mu^x}{x!}$$

Ejemplo 3.39 Sea X el número de criaturas de un tipo particular capturadas en una trampa durante un lapso de tiempo dado. Suponga que X tiene una distribución de Poisson con $\mu = 4.5$, así que en promedio las trampas contendrán 4.5 criaturas. [El artículo "Dispersal Dynamics of the Bivalve *Gemma gemma* in a Patchy Environment" (*Ecological Monographs*, 1995: 1–20) sugiere este modelo: el molusco bivalvo *Gemma gemma* es una pequeña almeja.] La probabilidad de que una trampa contenga exactamente cinco criaturas es

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = .1708$$

La probabilidad de que una trampa contenga cuando mucho cinco criaturas es

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4.5}(4.5)^x}{x!} = e^{-4.5}\left[1 + 4.5 + \frac{(4.5)^2}{2!} + \cdots + \frac{(4.5)^5}{5!}\right] = .7029$$ ■

La distribución de Poisson como límite

La siguiente proposición suministra la razón de ser en el uso de la distribución de Poisson en muchas situaciones.

PROPOSICIÓN

Suponga que en la función de masa de probabilidad binomial $b(x; n, p)$, $n \to \infty$ y $p \to 0$ de tal modo que np tienda a un valor $\mu > 0$. Entonces $b(x; n, p) \to p(x; \mu)$.

De acuerdo con esta proposición, *en cualquier experimento binomial en el cual n es grande y p es pequeña*, $b(x; n, p) \approx p(x; \mu)$, donde $\mu = np$. Como regla empírica, esta aproximación puede ser aplicada con seguridad si $n > 50$ y $np < 5$.

Ejemplo 3.40 Si un editor de libros no técnicos hace todo lo posible por que sus libros estén libres de errores tipográficos, de modo que la probabilidad de que cualquier página dada contenga por lo menos uno de esos errores es de .005 y los errores son independientes de una página a otra, ¿cuál es la probabilidad de que una de sus novelas de 400 páginas contenga exactamente una página con errores? ¿Cuándo mucho tres páginas con errores?

Con S denotando una página que contiene por lo menos un error y F una página libre de errores, el número X de páginas que contienen por lo menos un error es una variable aleatoria binomial con $n = 400$ y $p = .005$, así que $np = 2$. Se desea

$$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = .270671$$

El valor binomial es $b(1; 400, .005) = .270669$, así que la aproximación es muy buena.

Asimismo,

$$\begin{aligned} P(X \leq 3) &\approx \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} e^{-2}\frac{2^x}{x!} \\ &= .135335 + .270671 + .270671 + .180447 \\ &= .8571 \end{aligned}$$

y éste de nuevo se aproxima bastante al valor binomial $P(X \leq 3) = .8576$. ■

La tabla 3.2 muestra la distribución de Poisson con $\mu = 3$ junto con tres distribuciones binomiales con $np = 3$ y la figura 3.8 (generada por S-Plus) ilustra una gráfica de la distribución de Poisson junto con las dos primeras distribuciones binomiales. La aproximación es de uso limitado con $n = 30$, pero desde luego la precisión es mejor con $n = 100$ y mucho mejor con $n = 300$.

Tabla 3.2 Comparación de la distribución de Poisson y tres distribuciones binomiales

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391	0.047553	0.049041	0.049787
1	0.141304	0.147070	0.148609	0.149361
2	0.227656	0.225153	0.224414	0.224042
3	0.236088	0.227474	0.225170	0.224042
4	0.177066	0.170606	0.168877	0.168031
5	0.102305	0.101308	0.100985	0.100819
6	0.047363	0.049610	0.050153	0.050409
7	0.018043	0.020604	0.021277	0.021604
8	0.005764	0.007408	0.007871	0.008102
9	0.001565	0.002342	0.002580	0.002701
10	0.000365	0.000659	0.000758	0.000810

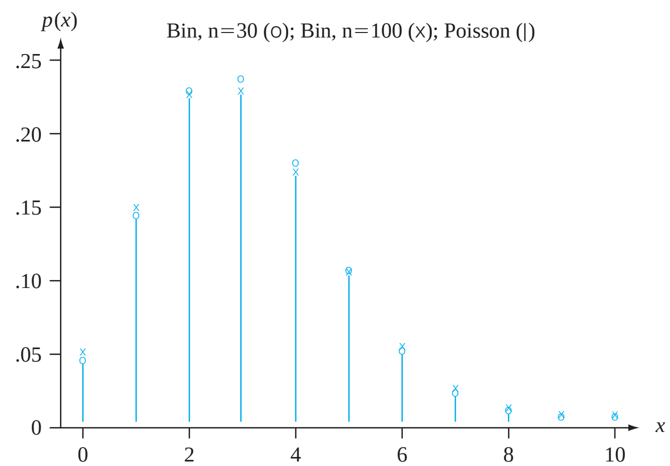


Figura 3.8 Comparación entre una distribución de Poisson y dos distribuciones binomiales

La tabla A.2 del apéndice muestra la función de distribución acumulativa $F(x; \mu)$ para $\mu = .1, .2, \ldots, 1, 2, \ldots, 10, 15$ y 20. Por ejemplo, si $\mu = 2$, entonces $P(X \leq 3) = F(3; 2) = .857$ como en el ejemplo 3.40, en tanto que $P(X = 3) = F(3; 2) - F(2; 2) = .180$. Alternativamente, muchos paquetes de computadora estadísticos generarán $p(x; \mu)$ y $F(x; \mu)$ al solicitarlo.

Media y varianza de X

Como $b(x; n, p) \to p(x; \mu)$ a medida que $n \to \infty$, $p \to 0$, $np \to \mu$, la media y la varianza de una variable binomial deberán aproximarse a las de una variable de Poisson. Estos límites son $np \to \mu$ y $np(1 - p) \to \mu$.

PROPOSICIÓN

Si X tiene una distribución de Poisson con parámetro μ, entonces $E(X) = V(X) = \mu$.

Estos resultados también pueden obtenerse directamente de la definición de media y varianza.

Ejemplo 3.41 (Continuación del ejemplo 3.39)

Tanto el número esperado de criaturas atrapadas como la varianza de éste son iguales a 4.5, y $\sigma_X = \sqrt{\mu} = \sqrt{4.5} = 2.12$. ■

Proceso de Poisson

Una aplicación muy importante de la distribución de Poisson surge en conexión con la ocurrencia de eventos de algún tipo en el transcurso del tiempo. Eventos de interés podrían ser visitas a un sitio web particular, pulsos de alguna clase registrados por un contador, mensajes de correo electrónico enviados a una dirección particular, accidentes en una instalación industrial o lluvias de rayos cósmicos observados por astrónomos en un observatorio particular. Se hace la siguiente suposición sobre la forma en que los eventos de interés ocurren:

1. Existe un parámetro $\alpha > 0$ tal que durante cualquier intervalo de tiempo corto Δt, la probabilidad de que ocurra exactamente un evento es $\alpha \cdot \Delta t + o(\Delta t)$*
2. La probabilidad de que ocurra más de un evento durante Δt es $o(\Delta t)$ [la que junto con la suposición 1, implica que la probabilidad de ningún evento durante Δt es $1 - \alpha \cdot \Delta t - o(\Delta t)$.
3. El número de eventos ocurridos durante este intervalo de tiempo Δt es independiente del número ocurrido antes de este intervalo de tiempo.

Informalmente, la suposición 1 dice que durante un corto intervalo de tiempo, la probabilidad de que ocurra un solo evento es aproximadamente proporcional a la duración del intervalo de tiempo, donde α es la constante de proporcionalidad. Ahora sea $P_k(t)$ la probabilidad de que k eventos serán observados durante cualquier intervalo de tiempo particular de duración t.

PROPOSICIÓN

$P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k/k!$, de modo que el número de eventos durante un intervalo de tiempo de duración t es una variable aleatoria de Poisson con parámetro $\mu = \alpha t$. El número esperado de eventos durante cualquier intervalo de tiempo es entonces αt, así que el número esperado durante un intervalo de tiempo unitario es α.

La ocurrencia de eventos en el transcurso del tiempo como se describió se llama *proceso de Poisson*; el parámetro α especifica *la rapidez* del proceso.

Ejemplo 3.42 Suponga que llegan pulsos a un contador a un ritmo promedio de seis por minuto, así que $\alpha = 6$. Para determinar la probabilidad de que en un intervalo de .5 minuto se reciba por lo menos un pulso, obsérvese que el número de pulsos en ese intervalo tiene una distribución de Poisson con parámetro $\alpha t = 6(.5) = 3$ (se utiliza .5 minuto porque α está expresada como rapidez por minuto). Entonces con $X =$ el número de pulsos recibidos en el intervalo de 30 segundos,

$$P(1 \le X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = .950$$ ■

En lugar de observar eventos en el transcurso del tiempo, considere observar eventos de algún tipo que ocurren en una región de dos o tres dimensiones. Por ejemplo, se podría seleccionar un mapa de una región R de un bosque, ir a dicha región y contar el número de árboles. Cada árbol representaría un evento que ocurre en un punto particular del espacio. Conforme a suposiciones similares a 1–3, se puede demostrar que el número de eventos que ocurren en una región R tiene una distribución de Poisson con parámetro $\alpha \cdot a(R)$, donde $a(R)$ es el área de R. La cantidad α es el número esperado de eventos por unidad de área o volumen.

* Una cantidad es $o(\Delta t)$ (léase "*o* pequeña de delta *t*") si, a medida que t tiende a cero, también lo hace $o(\Delta t)/\Delta t$. Es decir, $o(\Delta t)$ es incluso más insignificante (tiende a 0 más rápido) que Δt mismo. La cantidad $(\Delta t)^2$ tiene esta propiedad, pero sen(Δt) no.

EJERCICIOS Sección 3.6 (79–93)

79. Sea X el número de imperfecciones superficiales de una caldera seleccionada al azar de un tipo que tiene una distribución de Poisson con parámetro $\mu = 5$. Use la tabla A.2 del apéndice para calcular las siguientes probabilidades
a. $P(X \leq 8)$ **b.** $P(X = 8)$ **c.** $P(9 \leq X)$
d. $P(5 \leq X \leq 8)$ **e.** $P(5 < X < 8)$

80. Sea X el número de anomalías que ocurren en el material de una región particular de un disco de turbina de gas en aviones. El artículo "Methodology for Probabilistic Life Prediction of Multiple-Anomaly Materials" (*Amer. Inst. of Aeronautics and Astronautics J.*, 2006: 787–793) propone una distribución de Poisson para X. Supongamos que $\mu = 4$.
a. Calcule $P(X \leq 4)$ y $P(X < 4)$.
b. Calcule $P(4 \leq X \leq 8)$.
c. Calcule $P(8 \leq X)$.
d. ¿Cuál es la probabilidad de que el número observado de anomalías sobrepase suvalor medio por no más de una desviación estándar?

81. Suponga que el número de conductores que viajan entre un origen y destino particulares durante un lapso de tiempo designado tiene una distribución de Poisson con parámetro $\mu = 20$ (sugerido en el artículo "Dynamic Ride Sharing: Theory and Practice", *J. of Transp. Engr.*, 1997:308–312). ¿Cuál es la probabilidad de que el número de conductores
a. será cuando mucho de 10?
b. será de más de 20?
c. será de entre 10 y 20, inclusive? Será estrictamente de entre 10 y 20?
d. estará dentro de 2 desviaciones estándar del valor medio?

82. Considere escribir en un disco de computadora y luego enviarlo a través de un certificador que cuenta el número de pulsos faltantes. Suponga que este número X tiene una distribución de Poisson con parámetro $\mu = .2$. (Sugerido en "Average Sample Number for Semi-Curtailed Sampling Using the Poisson Distribution", *J. Quality Technology*, 1983: 126–129.)
a. ¿Cuál es la probabilidad de que un disco tenga exactamente un pulso faltante?
b. ¿Cuál es la probabilidad de que un disco tenga por lo menos dos pulsos faltantes?
c. Si se seleccionan dos discos independientemente, ¿cuál es la probabilidad de que ninguno contenga un pulso faltante?

83. Un artículo en *Los Ángeles Times* (3 de diciembre de 1993) reporta que 1 de cada 200 personas porta el gen defectuoso que provoca cáncer de colon hereditario. En una muestra de 1000 individuos, ¿cuál es la distribución aproximada del número que porta este gen? Use esta distribución para calcular la probabilidad aproximada de que
a. Entre 5 y 8 (inclusive) porten el gen.
b. Por lo menos 8 porten el gen.

84. Suponga que sólo .10% de todas las computadoras de cierto tipo experimentan fallas del CPU durante el periodo de garantía. Considere una muestra de 10,000 computadoras.
a. ¿Cuáles son el valor esperado y la desviación estándar del número de computadoras en la muestra que tienen el defecto?
b. ¿Cuál es la probabilidad (aproximada) de que más de 10 computadoras muestreadas tengan el defecto?
c. ¿Cuál es la probabilidad (aproximada) de que ninguna computadora muestreada tenga el defecto?

85. Suponga que una pequeña aeronave aterriza en un aeropuerto de acuerdo con un proceso de Poisson con razón $\alpha = 8$ por hora, de modo que el número de aterrizajes durante un lapso de tiempo de t horas es una variable aleatoria de Poisson con parámetro $\mu = 8t$.
a. ¿Cuál es la probabilidad de que exactamente 6 aeronaves pequeñas aterricen durante un intervalo de 1 hora? ¿Por lo menos 6? ¿Por lo menos 10?
b. ¿Cuáles son el valor esperado y la desviación estándar del número de aeronaves pequeñas que aterrizan durante un lapso de 90 min?
c. ¿Cuál es la probabilidad de que por lo menos 20 aeronaves pequeñas aterricen durante un lapso de 2.5 horas? ¿De qué cuando mucho aterricen 10 durante este periodo?

86. El número de personas que llegan para tratamiento a una sala de urgencias puede ser modelado mediante un proceso de Poisson con parámetro de rapidez de cinco por hora.
a. ¿Cuál es la probabilidad de que ocurran exactamente cuatro arribos durante una hora particular?
b. ¿Cuál es la probabilidad de que por lo menos cuatro personas arriben durante una hora particular?
c. ¿Cuántas personas espera que arriben durante un periodo de 45 min?

87. El número de solicitudes de ayuda recibidas por un servicio de grúas es un proceso de Poisson con razón $\alpha = 4$ por hora.
a. Calcule la probabilidad de que exactamente diez solicitudes sean recibidas durante un periodo particular de 2 horas.
b. Si los operadores del servicio de grúas hacen una pausa de 30 minutos para el almuerzo, ¿cuál es la probabilidad de que no dejen de atender llamadas de ayuda?
c. ¿Cuántas llamadas esperaría durante esta pausa?

88. Al someter a prueba tarjetas de circuito, la probabilidad de que cualquier diodo particular falle es de .01. Suponga que una tarjeta de circuito contiene 200 diodos.
a. ¿Cuántos diodos esperaría que fallen y cuál es la desviación estándar del número que se espera fallen?
b. ¿Cuál es la probabilidad (aproximada) de que por lo menos cuatro diodos fallen en una tarjeta seleccionada al azar?
c. Si se envían cinco tarjetas a un cliente particular, ¿qué tan probable es que por lo menos cuatro de ellas funcionen apropiadamente? (Una tarjeta funciona apropiadamente sólo si todos sus diodos funcionan.)

89. El artículo "Reliability-Based Service-Life Assessment of Aging Concrete Structures" (*J. Structural Engr.*, 1993: 1600–1621) sugiere que un proceso de Poisson puede ser utilizado para representar la ocurrencia de cargas estructurales en el transcurso del tiempo. Suponga que el tiempo medio entre ocurrencias de cargas es de .5 al año.
a. ¿Cuántas cargas se espera que ocurran durante un periodo de 2 años?

b. ¿Cuál es la probabilidad de que ocurran más de cinco cargas durante un periodo de 2 años?

c. ¿Qué tan largo debe ser un periodo de modo que la probabilidad de que no ocurran cargas durante dicho periodo sea cuando mucho de .1?

90. Sea X que tiene una distribución de Poisson con parámetro μ. Demuestre que $E(X) = \mu$ derivada directamente de la definición de valor esperado. [*Sugerencia*: el primer término en la suma es igual a 0 y luego x puede ser eliminada. Ahora saque como factor a μ y demuestre que lo que queda suma 1.]

91. Suponga que hay árboles distribuidos en un bosque de acuerdo con un proceso de Poisson bidimensional con parámetro α, el número esperado de árboles por acre es de 80.

a. ¿Cuál es la probabilidad de que en un terreno de cuarto de acre, habrá cuando mucho 16 árboles?

b. Si el bosque abarca 85,000 acres, ¿cuál es el número esperado de árboles en el bosque?

c. Suponga que selecciona un punto en el bosque y construye un círculo de .1 milla de radio. Sea X = el número de árboles dentro de esa región circular. ¿Cuál es la función de masa de probabilidad de X? [*Sugerencia*: 1 milla cuadrada = 640 acres.]

92. A una estación de inspección de equipo vehicular llegan automóviles de acuerdo con un proceso de Poisson con razón $\alpha = 10$ por hora. Suponga que un vehículo que llega con probabilidad de .5 no tendrá violaciones de equipo.

a. ¿Cuál es la probabilidad de que exactamente diez lleguen durante la hora y que ninguno tenga violaciones?

b. Con cualquier $y \geq 10$ fija, ¿cuál es la probabilidad de que y lleguen durante la hora, diez de los cuales no tengan violaciones?

c. ¿Cuál es la probabilidad de que lleguen diez carros "sin violaciones" durante la siguiente hora? [*Sugerencia*: sume las probabilidades en el inciso (b) desde $y = 10$ hasta ∞.]

93. **a.** En un proceso de Poisson, ¿qué tiene que suceder tanto en el intervalo de tiempo $(0, t)$ como en el intervalo $(t, t + \Delta t)$ de modo que no ocurran eventos en todo el intervalo $(0, t + \Delta t)$? Use esto y las suposiciones 1–3 para escribir una relación entre $P_0(t + \Delta t)$ y $P_0(t)$.

b. Use el resultado del inciso (a) para escribir una expresión para la diferencia $P_0(t + \Delta t) - P_0(t)$. Divida entonces entre Δt y permita que $\Delta t \to 0$ para obtener una ecuación que implique $(d/dt)P_0(t)$, la derivada de $P_0(t)$ con respecto a t.

c. Verifique que $P_0(t) = e^{-\alpha t}$ satisface la ecuación del inciso (b).

d. Se puede demostrar de manera similar a los incisos (a) y (b) que los $P_k(t)$ deben satisfacer el sistema de ecuaciones diferenciales

$$\frac{d}{dt}P_k(t) = \alpha P_{k-1}(t) - \alpha P_k(t)$$
$$k = 1, 2, 3, \ldots$$

Verifique que $P_k(t) = e^{-\alpha t}(\alpha t)^k/k!$ satisface el sistema. (En realidad ésta es la única solución.)

EJERCICIOS SUPLEMENTARIOS (94–122)

94. Considere un mazo compuesto de siete cartas, marcadas 1, 2, . . . , 7. Se seleccionan al azar tres de estas cartas. Defina una variable aleatoria W como W = la suma de los números resultantes y calcule la función de masa de probabilidad de W. Calcule entonces μ y σ^2. [*Sugerencia*: considere los resultados sin orden, de modo que (1, 3, 7) y (3, 1, 7) no son resultados diferentes. Entonces existen 35 resultados y pueden ser puestos en lista. (Este tipo de variable aleatoria en realidad se presenta en conexión con una prueba de hipótesis llamada prueba de suma de renglones de Wilcoxon, en la cual hay una muestra x y una muestra y y W es la suma de los renglones de x en la muestra combinada; véase la sección 15.2.)

95. Después de barajar un mazo de 52 cartas, un tallador reparte 5. Sea X = el número de palos representados en la mano de 5 cartas.

a. Demuestre que la función de masa de probabilidad de X es

x	1	2	3	4
$p(x)$	.002	.146	.588	.264

[*Sugerencia*: $p(1) = 4P$(todas son espadas), $p(2) = 6P$(sólo espadas y corazones con por lo menos una de cada palo) y $p(4) = 4P$(2 espadas $\cap$ una de cada otro palo).]

b. Calcule μ, σ^2 y σ.

96. La variable aleatoria binomial negativa X se definió como el número de fallas (F) que preceden al éxito (S) r-ésimo. Sea Y = el número de ensayos necesarios para obtener el *éxito* (S) r-ésimo. Del mismo modo en que fue obtenida la función de masa de probabilidad de X, deduzca la función de masa de probabilidad de Y.

97. De todos los clientes que adquieren abrepuertas de cochera automáticos, 75% adquieren el modelo de transmisión por cadena. Sea X = el número entre los siguientes 15 compradores que seleccionan el modelo de transmisión por cadena.

a. ¿Cuál es la función de masa de probabilidad de X?

b. Calcule $P(X > 10)$.

c. Calcule $P(6 \leq X \leq 10)$.

d. Calcule μ y σ^2.

e. Si la tienda actualmente tiene en existencia 10 modelos de transmisión por cadena y 8 modelos de transmisión por flecha, ¿cuál es la probabilidad de que las solicitudes de estos 15 clientes puedan ser satisfechas con las existencias actuales?

98. Un amigo recientemente planeó un viaje de campamento. Tenía dos linternas, una que requería una sola batería de 6 V y otra que utilizaba dos baterías de tamaño D. Antes había empacado dos baterías de 6 V y cuatro tamaño D en su camper. Suponga que la probabilidad de que cualquier batería particular funcione es p y que las baterías funcionan o fallan independientemente una de otra. Nuestro amigo desea llevar sólo una linterna. ¿Con qué valores de p deberá llevar la linterna de 6 V?

99. Un *sistema k de n* es uno que funcionará si y sólo si por lo menos k de los n componentes individuales en el sistema funcionan. Si los componentes individuales funcionan independientemente uno de otro, cada uno con probabilidad de .9, ¿cuál es la probabilidad de que un sistema 3 de 5 funcione?

100. Un fabricante de chips de circuitos integrados desea controlar la calidad de sus productos rechazando cualquier lote en el que la proporción de chips sea demasiado alta. Con esta finalidad, de cada lote de 10,000 chips, se seleccionarán y probarán 25. Si por lo menos 5 de éstos están defectuosos, todo el lote será rechazado.

a. ¿Cuál es la probabilidad de que un lote será rechazado si 5% de los chips en el lote están de hecho, defectuosos?

b. Responda la pregunta del inciso (a) si el porcentaje de chips defectuosos es 10%.

c. Responda la pregunta del inciso (a) si el porcentaje de chips defectuosos es 20%.

d. ¿Qué les sucedería a las probabilidades en los incisos (a)–(c) si el número de rechazo crítico se incrementara de 5 a 6?

101. De las personas que pasan a través de un detector de metales en un aeropuerto, el .5% lo activan; sea $X =$ el número entre un grupo de 500 seleccionado al azar que activan el detector.

a. ¿Cuál es la función de masa de probabilidad (aproximada) de X?

b. Calcule $P(X = 5)$.

c. Calcule $P(5 \leq X)$.

102. Una firma consultora educativa está tratando de decidir si los estudiantes de preparatoria que nunca antes han utilizado una calculadora de mano pueden resolver cierto tipo de problema más fácilmente con una calculadora que utiliza lógica polaca inversa o una que no utiliza esta lógica. Se selecciona una muestra de 25 estudiantes y se les permite practicar con ambas calculadoras. Luego a cada estudiante se le pide que resuelva un problema con la calculadora polaca inversa y un problema similar con la otra. Sea $p = P(S)$, donde S indica que un estudiante resolvió el problema más rápido con la lógica polaca inversa que sin ella y sea $X =$ número de éxitos.

a. Si $p = .5$, ¿cuál es $P(7 \leq X \leq 18)$?

b. Si $p = .8$, ¿cuál es $P(7 \leq X \leq 18)$?

c. Si la pretensión de que $p = .5$ tiene que ser rechazada cuando $x \leq 7$ o $x \geq 18$, ¿cuál es la probabilidad de rechazar la pretensión cuando en realidad es correcta?

d. Si la decisión de rechazar la pretensión $p = .5$ se hace como en el inciso (c), ¿cuál es la probabilidad de que la pretensión no sea rechazada cuando $p = .6$? ¿Cuándo $p = .8$?

e. ¿Qué regla de decisión escogería para rechazar la pretensión de que $p = .5$ si desea que la probabilidad en el inciso (c) sea cuando mucho de .01?

103. Considere una enfermedad cuya presencia puede ser identificada por medio de un análisis de sangre. Sea p la probabilidad de que un individuo seleccionado al azar tenga la enfermedad. Suponga que se seleccionan independientemente n individuos para analizarlos. Una forma de proceder es analizar cada una de las n muestras de sangre. Un procedimiento potencialmente más económico, de análisis en grupo, se introdujo durante la Segunda Guerra Mundial para identificar hombres sifilíticos entre los reclutas. En primer lugar, se toma una parte de cada muestra de sangre, se combinan estos especímenes y se realiza un solo análisis. Si ninguno tiene la enfermedad, el resultado será negativo y sólo se requiere un análisis. Si por lo menos un individuo está enfermo, el análisis de la muestra combinada dará un resultado positivo, en cuyo caso se realizan los análisis de los n individuos. Si $p = .1$ y $n = 3$, ¿cuál es el número esperado de análisis si se utiliza este procedimiento? ¿Cuál es el número esperado cuando $n = 5$? [El artículo "Random Multiple-Access Communication and Group Testing" (*IEEE Trans. on Commun.*, 1984: 769–774) aplicó estas ideas a un sistema de comunicación en el cual la dicotomía fue usuario ocioso/activo en lugar de enfermo/no enfermo.]

104. Sea p_1 la probabilidad de que cualquier símbolo de código particular sea erróneamente transmitido a través de un sistema de comunicación. Suponga que en diferentes símbolos, ocurren errores de manera independiente uno de otro. Suponga también que con probabilidad p_2 un símbolo erróneo es corregido al ser recibido. Sea X el número de símbolos correctos en un bloque de mensaje compuesto de n símbolos (una vez que el proceso de corrección ha terminado). ¿Cuál es la distribución de probabilidad de X?

105. El comprador de una unidad generadora de potencia requiere de c arranques consecutivos exitosos antes de aceptar la unidad. Suponga que los resultados de arranques individuales son independientes entre sí. Sea p la probabilidad de que cualquier arranque particular sea exitoso. La variable aleatoria de interés es $X =$ el número de arranques que deben hacerse antes de la aceptación. Dé la función de masa de probabilidad de X en el caso $c = 2$. Si $p = .9$, ¿cuál es $P(X \leq 8)$? [*Sugerencia*: con $x \geq 5$, exprese $p(x)$ "recursivamente" en función de la función de masa de probabilidad evaluada con los valores más pequeños $x - 3, x - 4, \ldots, 2$.] (Este problema fue sugerido del artículo "Evaluation of a Start-Up Demonstration Test", *J. Quality Technology*, 1983: 103–106.)

106. Una aerolínea ha desarrollado un plan para un club de viajeros ejecutivos sobre la premisa de que 10% de sus clientes actuales calificarían para la membresía.

a. Suponiendo la validez de esta premisa, entre 25 clientes actuales seleccionados al azar, ¿cuál es la probabilidad de que entre 2 y 6 (inclusive) califiquen para la membresía?

b. De nuevo suponiendo la validez de la premisa, ¿cuál es el número esperado de clientes que califican y la desviación estándar del número que califica en una muestra aleatoria de 100 clientes actuales?

c. Sea X el número en una muestra al azar de 25 clientes actuales que califican para la membresía. Considere rechazar la premisa de la compañía a favor de la pretensión de que $p > .10$ si $x \geq 7$. ¿Cuál es la probabilidad de que la premisa de la compañía sea rechazada cuando en realidad es válida?

d. Remítase a la regla de decisión introducida en el inciso (c). ¿Cuál es la probabilidad de que la premisa de la compañía no sea rechazada aun cuando $p = .20$ (es decir, 20% califican)?

107. Cuarenta por ciento de las semillas de mazorcas de maíz (maíz moderno) portan sólo una espiga y el 60% restante portan dos espigas. Una semilla con una espiga producirán una mazorca con espigas únicas 29% del tiempo, en tanto que una semilla con dos espigas producirán una mazorca con espigas únicas 26% del tiempo. Considere seleccionar al azar diez semillas.

a. ¿Cuál es la probabilidad de que exactamente cinco de estas semillas porten una sola espiga y de que produzcan una mazorca con una sola espiga?

b. ¿Cuál es la probabilidad de que exactamente cinco de estas mazorcas producidas por estas semillas tengan espigas únicas? ¿Cuál es la probabilidad de que cuando mucho cinco mazorcas tengan espigas únicas?

108. Un juicio terminó con el jurado en desacuerdo porque ocho de sus miembros estuvieron a favor de un veredicto de culpabilidad y los otros cuatro estuvieron a favor de la absolución. Si los jurados salen de la sala en orden aleatorio y cada uno de los primeros cuatro que salen de la sala es acosado por un reportero para entrevistarlo, ¿cuál es la función de masa de probabilidad de X = el número de jurados a favor de la absolución entre los entrevistados? ¿Cuántos de los que están a favor de la absolución espera que sean entrevistados?

109. Un servicio de reservaciones emplea cinco operadores de información que reciben solicitudes de información independientemente uno de otro, cada uno de acuerdo con un proceso de Poisson con rapidez $\alpha = 2$ por minuto.

a. ¿Cuál es la probabilidad de que durante un periodo de 1 minuto dado, el primer operador no reciba solicitudes?

b. ¿Cuál es la probabilidad de que durante un periodo de 1 minuto dado, exactamente cuatro de los cinco operadores no reciban solicitudes?

c. Escriba una expresión para la probabilidad de que durante un periodo de 1 minuto dado, todos los operadores reciban exactamente el mismo número de solicitudes.

110. En un gran campo se distribuyen al azar las langostas de acuerdo con una distribución de Poisson con parámetro $\alpha = 2$ por yarda cuadrada. ¿Qué tan grande deberá ser el radio R de una región de muestreo circular para que la probabilidad de hallar por lo menos una en la región sea igual a .99?

111. Un puesto de periódicos ha pedido cinco ejemplares de cierto número de una revista de fotografía. Sea X = el número de individuos que vienen a comprar esta revista. Si X tiene una distribución de Poisson con parámetro $\mu = 4$, ¿cuál es el número esperado de ejemplares que serán vendidos?

112. Los individuos A y B comienzan a jugar una secuencia de partidas de ajedrez. Sea $S = \{\text{A gana un juego}\}$ y suponga que los resultados de juegos sucesivos son independientes con $P(S) = p$ y $P(F) = 1 - p$ (nunca empatan). Jugarán hasta que uno de ellos gane diez juegos. Sea X = el número de partidas jugadas (con posibles valores 10, 11, . . . , 19).

a. Con $x = 10, 11, \ldots, 19$, obtenga una expresión para $p(x) = P(X = x)$.

b. Si un empate es posible, con $p = P(S)$, $q = P(F)$, $1 - p - q = P(\text{empate})$, ¿cuáles son los posibles valores de X? ¿Cuál es $P(20 \le X)$? [*Sugerencia*: $P(20 \le X) = 1 - P(X < 20)$.]

113. Un análisis para detectar la presencia de una enfermedad tiene una probabilidad de .20 de dar un resultado falso positivo (que indica que un individuo tiene la enfermedad cuando éste no es el caso) y una probabilidad de .10 de dar un resultado falso negativo. Suponga que diez individuos son analizados, cinco de los cuales tienen la enfermedad y cinco de los cuales no. Sea X = el número de lecturas positivas que resultan.

a. ¿Tiene X una distribución binomial? Explique su razonamiento.

b. ¿Cuál es la probabilidad de que exactamente tres de diez resultados sean positivos?

114. La función de masa de probabilidad binomial negativa generalizada está dada por

$$nb(x; r, p) = k(r, x) \cdot p^r(1 - p)^x$$
$$x = 0, 1, 2, \ldots$$

Sea X el número de plantas de cierta especie encontradas en una región particular y tenga esta distribución con $p = .3$ y $r = 2.5$. ¿Cuál es $P(X = 4)$? ¿Cuál es la probabilidad de que por lo menos se encuentre una planta?

115. Hay dos contadores públicos en una oficina particular que preparan declaraciones de impuestos para los clientes. Supongamos que para un tipo particular de forma compleja, el número de errores cometidos por el preparador de la primera tiene una distribución de Poisson con media μ_1, el número de errores cometidos por el preparador de la segunda tiene una distribución de Poisson con media μ_2 y que cada contador prepara el mismo número de formas de este tipo. Entonces, si una forma de este tipo es seleccionada al azar, la función

$$p(x; \mu_1, \mu_2) = .5\frac{e^{-\mu_1}\mu_1^x}{x!} + .5\frac{e^{-\mu_2}\mu_2^x}{x!} \quad x = 0, 1, 2, \ldots$$

da la función de masa de probabilidad de X = el número de errores en el formulario seleccionado.

a. Compruebe que $p(x; \mu_1, \mu_2)$ es de hecho una función de masa de probabilidad legítima (≥ 0 y se suma a 1).

b. ¿Cuál es el número esperado de errores en el formulario seleccionado?

c. ¿Cuál es la varianza del número de errores en el formulario seleccionado?

d. ¿Cómo cambia la función de masa de probabilidad si el primer contador prepara el 60% de todas esas formas y el segundo prepara el 40%?

116. La *moda* de una variable aleatoria discreta X con función de masa de probabilidad $p(x)$ es ese valor x^* con el cual $p(x)$ alcanza su valor más grande (el valor x más probable).

a. Sea $X \sim \text{Bin}(n, p)$. Considerando la razón $b(x + 1; n, p)/b(x; n, p)$, demuestre que $b(x; n, p)$ se incrementa con x en tanto $x < np - (1 - p)$. Concluya que la moda x^* es el entero que satisface $(n + 1)p - 1 \le x^* \le (n + 1)p$.

b. Demuestre que si X tiene una distribución de Poisson con parámetro μ, la moda es el entero más grande menor que μ. Si μ es un entero, demuestre que tanto $\mu - 1$ como μ son modas.

117. Un disco duro de computadora tiene diez pistas concéntricas, numeradas 1, 2, . . . , 10 desde la más externa hasta la más interna y un solo brazo de acceso. Sea p_i = la probabilidad de que cualquier solicitud particular de datos hará que el brazo se vaya a la pista $i (i = 1, \ldots, 10)$. Suponga que las pistas recorridas en búsquedas sucesivas son independientes. Sea X = el número de pistas sobre las cuales pasa el brazo de acceso durante dos solicitudes sucesivas (excluida la pista que el brazo acaba de dejar, así que los valores posibles son $x = 0$, 1, . . . , 9). Calcule la función de masa de probabilidad de X. [*Sugerencia*: P(el brazo está ahora sobre la pista i y $X = j$) $=$ $P(X = j|$el brazo está ahora sobre $i) \cdot p_i$. Una vez que se escribe la probabilidad condicional en función de $p_1, \ldots, p_{10}$, mediante la ley de probabilidad total, se obtiene la probabilidad deseada sumando a lo largo de i.]

118. Si X es una variable aleatoria hipergeométrica demuestre directamente con la definición que $E(X) = nM/N$ (considere sólo el caso $n < M$). [*Sugerencia*: saque como factor a nM/N de la suma para $E(X)$ y demuestre que los términos adentro de la suma son de la forma $h(y; n - 1, M - 1, N - 1)$ donde $y = x - 1$.]

119. Use el hecho de que

$$\sum_{\text{toda } x} (x - \mu)^2 p(x) \geq \sum_{x:\, |x-\mu| \geq k\sigma} (x - \mu)^2 p(x)$$

para comprobar la desigualdad de Chebyshev dada en el ejercicio 44.

120. El proceso de Poisson simple de la sección 3.6 está caracterizado por una rapidez constante α a la cual los eventos ocurren por unidad de tiempo. Una generalización de esto es suponer que la probabilidad de que ocurra exactamente un evento en el intervalo $[t, t + \Delta t]$ es $\alpha(t) \cdot \Delta t + o(\Delta t)$. Se puede demostrar entonces que el número de eventos que ocurren durante un intervalo $[t_1, t_2]$ tiene una distribución de Poisson con parámetro

$$\mu = \int_{t_2}^{t_1} \alpha(t)\, dt$$

La ocurrencia de eventos en el transcurso del tiempo en esta situación se llama *proceso de Poisson no homogéneo*. El artículo "Inference Based on Retrospective Ascertainment", *J. Amer. Stat. Assoc.*, 1989: 360–372, considera la función de intensidad

$$\alpha(t) = e^{a+bt}$$

apropiada para eventos que implican la transmisión de VIH (el virus del SIDA) vía transfusiones sanguíneas. Suponga que $a = 2$ y $b = .6$ (cercanos a los valores sugeridos en el artículo), con el tiempo en años.

a. ¿Cuál es el número esperado de eventos en el intervalo [0, 4]? ¿En [2, 6]?

b. ¿Cuál es la probabilidad de que cuando mucho ocurran 15 eventos en el intervalo [0, .9907]?

121. Considere un conjunto de $A_1, \ldots, A_k$ de eventos mutuamente exclusivos y exhaustivos y una variable aleatoria X cuya distribución depende de cuál de los eventos A_i ocurra (p. ej., un viajero podría seleccionar una de tres rutas posibles de su casa al trabajo, con X como el tiempo de recorrido). Sea $E(X|A_i)$ el valor esperado de X dado que el evento A_i ocurre. Entonces se puede demostrar que $E(X) = \Sigma E(X|A_i) \cdot P(A_i)$ es el promedio ponderado de las "expectativas condicionales" individuales donde las ponderaciones son las probabilidades de la división de eventos.

a. La duración esperada de una llamada de voz a un número telefónico particular es de 3 minutos, mientras que la duración esperada de una llamada de datos a ese mismo número es de 1 minuto. Si 75% de las llamadas son de voz, ¿cuál es la duración esperada de la siguiente llamada?

b. Una pastelería vende tres diferentes tipos de galletas con chispas de chocolate. El número de chispas de chocolate en un tipo i de galleta tiene una distribución de Poisson con parámetro $\mu_i = i + 1 \quad (i = 1, 2, 3)$. Si 20% de todos los clientes que compran una galleta con chispas de chocolate selecciona el primer tipo, 50% elige el segundo tipo y el 30% restante opta por el tercer tipo, ¿cuál es el número esperado de chispas en una galleta comprada por el siguiente cliente?

122. Considere una fuente de comunicaciones que transmite paquetes que contienen lenguaje digitalizado. Después de cada transmisión, el receptor envía un mensaje que indica si la transmisión fue exitosa o no. Si una transmisión no es exitosa, el paquete es reenviado. Suponga que el paquete de voz puede ser transmitido un máximo de 10 veces. Suponiendo que los resultados de transmisiones sucesivas son independientes uno de otro y que la probabilidad de que cualquier transmisión particular sea exitosa es p, determine la función de masa de probabilidad de la variable aleatoria $X =$ el número de veces que un paquete es transmitido. Luego obtenga una expresión para el número de veces que se espera que un paquete sea transmitido.

Bibliografía

Johnson, Norman, Samuel Kotz y Adrienne Kemp, *Discrete Univariate Distributions*. Wiley, Nueva York, 1992. Una enciclopedia de información sobre distribuciones discretas.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Contiene una discusión a fondo tanto de las propiedades generales de distribuciones discretas y continuas como los resultados para distribuciones específicas.

Ross, Sheldon, *Introduction to Probability Models* (9a. ed.), Academic Press, Nueva York, 2007. Una buena fuente de material sobre el proceso de Poisson y generalizaciones, y una amena introducción a otros temas de probabilidad aplicada.

4 Variables aleatorias continuas y distribuciones de probabilidad

INTRODUCCIÓN

El capítulo 3 se concentró en el desarrollo de distribuciones de probabilidad de variables aleatorias discretas. En este capítulo se estudia el segundo tipo general de variable aleatoria que se presenta en muchos problemas aplicados. Las secciones 4.1 y 4.2 presentan las definiciones y propiedades básicas de las variables aleatorias continuas y sus distribuciones de probabilidad. En la sección 4.3 se estudia con detalle la variable aleatoria normal y su distribución, sin duda la más importante y útil en la probabilidad y estadística. Las secciones 4.4 y 4.5 se ocupan de otras distribuciones continuas utilizadas con frecuencia en trabajo aplicado. En la sección 4.6 se introduce un método de evaluar si un dato muestral es compatible con una distribución especificada.

4.1 Funciones de densidad de probabilidad

Una variable aleatoria discreta es una cuyos valores posibles constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita (una lista en la cual existe un primer elemento, un segundo elemento, etc.). Una variable aleatoria cuyo conjunto de valores posibles es un intervalo completo de números no es discreta.

De acuerdo con el capítulo 3 recuérdese que una variable aleatoria X es continua si (1) sus valores posibles comprenden un solo intervalo sobre la línea de numeración (para alguna $A < B$, cualquier número x entre A y B es un valor posible) o una unión de intervalos disjuntos y (2) $P(X = c) = 0$ para cualquier número c que sea un valor posible de X.

Ejemplo 4.1 En el estudio de la ecología de un lago, se mide la profundidad en lugares seleccionados, entonces $X =$ la profundidad en ese lugar es una variable aleatoria continua. En este caso A es la profundidad mínima en la región muestreada y B es la profundidad máxima. ■

Ejemplo 4.2 Si se selecciona al azar un compuesto químico y se determina su pH X, entonces X es una variable aleatoria continua porque cualquier valor pH entre 0 y 14 es posible. Si se conoce más sobre el compuesto seleccionado para su análisis, entonces el conjunto de posibles valores podría ser un subintervalo de [0, 14], tal como $5.5 \leq x \leq 6.5$, pero X seguiría siendo continua. ■

Ejemplo 4.3 Sea X la cantidad de tiempo que un cliente seleccionado al azar pasa esperando antes de que comience su corte de pelo. El primer pensamiento podría ser que X es una variable aleatoria continua, puesto que se requiere medirla para determinar su valor. Sin embargo, existen clientes suficientemente afortunados que no tienen que esperar antes de sentarse en el sillón del peluquero. Así que el caso debe ser $P(X = 0) > 0$. Aunque, en caso de que no haya sillones vacíos, el tiempo de espera será continuo puesto que X podría asumir entonces cualquier valor entre un tiempo mínimo posible A y un tiempo máximo posible B. Esta variable aleatoria no es ni puramente discreta ni puramente continua sino que es una mezcla de los dos tipos. ■

Se podría argumentar que aunque en principio variables tales como altura, peso y temperatura son continuas, en la práctica las limitaciones de los instrumentos de medición nos restringen a un mundo discreto (aunque en ocasiones muy finamente subdividido). Sin embargo, los modelos continuos a menudo representan muy bien de forma aproximada situaciones del mundo real y con frecuencia es más fácil trabajar con matemáticas continuas (el cálculo) que con matemáticas de variables discretas y distribuciones.

Distribuciones de probabilidad de variables continuas

Supóngase que la variable X de interés es la profundidad de un lago en un punto sobre la superficie seleccionado al azar. Sea $M =$ la profundidad máxima (en metros), así que cualquier número en el intervalo $[0, M]$ es un valor posible de X. Si se "discretiza" X midiendo la profundidad al metro más cercano, entonces los valores posibles son enteros no negativos menores o iguales a M. La distribución discreta de profundidad resultante se ilustra con un histograma de probabilidad. Si se traza el histograma de modo que el área del rectángulo sobre cualquier entero posible k sea la proporción del lago cuya profundidad es (al metro más cercano) k, entonces el área total de todos los rectángulos es 1. En la figura 4.1(a) aparece un posible histograma.

Si se mide la profundidad con mucho más precisión y se utiliza el mismo eje de medición de la figura 4.1(a), cada rectángulo en el histograma de probabilidad resultante es mucho más angosto, aun cuando el área total de todos los rectángulos sigue siendo 1.

En la figura 4.1(b) se ilustra un posible histograma; tiene una apariencia mucho más regular que el histograma de la figura 4.1(a). Si se continúa de esta manera midiendo la profundidad más y más finamente, la secuencia resultante de histogramas se aproxima a una curva más regular, tal como la ilustrada en la figura 4.1(c). Como en cada histograma el área total de todos los rectángulos es igual a 1, el área total bajo la curva regular también es 1. La probabilidad de que la profundidad en un punto seleccionado al azar se encuentre entre a y b es simplemente el área bajo la curva regular entre a y b. Es de manera exacta una curva suave del tipo ilustrado en la figura 4.1(c) la que especifica una distribución de probabilidad continua.

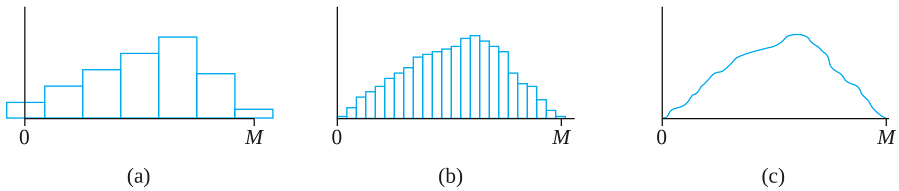


Figura 4.1 (a) Histograma de probabilidad de profundidad medida al metro más cercano; (b) histograma de probabilidad de profundidad medida al centímetro más cercano; (c) un límite de una secuencia de histogramas discretos

DEFINICIÓN

Sea X una variable aleatoria continua. Entonces, una **distribución de probabilidad** o **función de densidad de probabilidad** (fdp) de X es una función $f(x)$ tal que para dos números cualesquiera a y b con $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Es decir, la probabilidad de que X asuma un valor en el intervalo $[a, b]$ es el área sobre este intervalo y bajo la gráfica de la función de densidad, como se ilustra en la figura 4.2. La gráfica de $f(x)$ a menudo se conoce como *curva de densidad*.

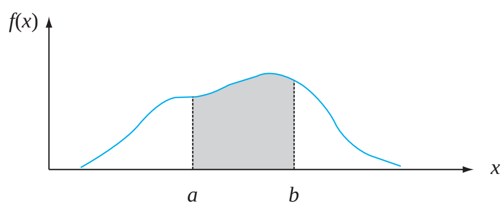


Figura 4.2 $P(a \leq X \leq b) =$ el área bajo la curva de densidad entre a y b

Para que $f(x)$ sea una función de densidad de probabilidad legítima, debe satisfacer las dos siguientes condiciones:

1. $f(x) \geq 0$ con todas las x

2. $\int_{-\infty}^{\infty} f(x)\,dx =$ área bajo toda la gráfica de $f(x)$

$= 1$

Ejemplo 4.4

La dirección de una imperfección con respecto a una línea de referencia sobre un objeto circular como un neumático, un rotor de freno o un volante está, en general, sujeta a incertidumbre. Considérese la línea de referencia que conecta el vástago de la válvula de un neumático con el punto central y sea X el ángulo medido en el sentido de las manecillas del reloj con respecto a la ubicación de una imperfección. Una posible función de densidad de probabilidad de X es

$$f(x) = \begin{cases} \dfrac{1}{360} & 0 \le x < 360 \\ 0 & \text{de lo contrario} \end{cases}$$

La función de densidad de probabilidad aparece graficada en la figura 4.3. Claramente $f(x) \ge 0$. El área bajo la curva de densidad es simplemente el área de un rectángulo: (altura)(base) $= \left(\frac{1}{360}\right)(360) = 1$. La probabilidad de que el ángulo sea de entre 90° y 180° es

$$P(90 \le X \le 180) = \int_{90}^{180} \frac{1}{360}\,dx = \frac{x}{360}\Bigg|_{x=90}^{x=180} = \frac{1}{4} = .25$$

La probabilidad de que el ángulo de ocurrencia esté dentro de 90° de la línea de referencia es

$$P(0 \le X \le 90) + P(270 \le X < 360) = .25 + .25 = .50$$

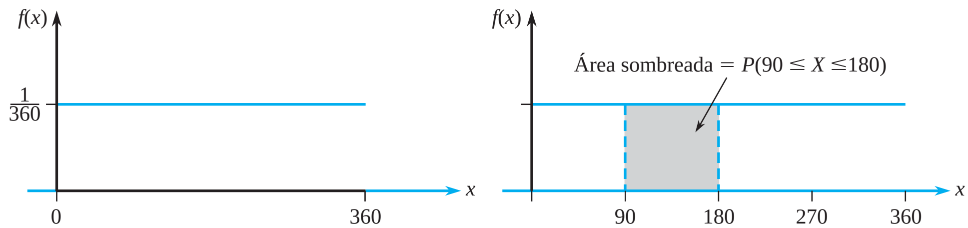


Figura 4.3 Función de densidad de probabilidad del ejemplo 4.4 ■

Como siempre que $0 \le a \le b \le 360$ en el ejemplo 4.4 y $P(a \le X \le b)$ depende sólo del ancho $b - a$ del intervalo, se dice que X tiene una distribución uniforme.

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución uniforme** en el intervalo $[A, B]$ si la función de densidad de probabilidad de X es

$$f(x; A, B) = \begin{cases} \dfrac{1}{B - A} & A \le x \le B \\ 0 & \text{de lo contrario} \end{cases}$$

La gráfica de cualquier función de densidad de probabilidad uniforme es como la de la figura 4.3, excepto que el intervalo de densidad positiva es $[A, B]$ en lugar de $[0, 360]$.

En el caso discreto, una función de masa de probabilidad (fmp) dice cuántas pequeñas "burbujas" de masa de probabilidad de varias magnitudes están distribuidas a lo largo del eje de medición. En el caso continuo, la densidad de probabilidad está "repartida" en forma continua a lo largo del intervalo de posibles valores. Cuando la densidad está distribuida uniformemente a lo largo del intervalo, se obtiene una función de densidad de probabilidad uniforme como en la figura 4.3.

Cuando X es una variable aleatoria discreta, a cada valor posible se le asigna una probabilidad positiva. Esto no es cierto en el caso de una variable aleatoria continua (es decir, se satisface la segunda condición de la definición) porque el área bajo una curva de densidad situada sobre cualquier valor único es cero:

$$P(X = c) = \int_c^c f(x)\,dx = \lim_{\varepsilon \to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)\,dx = 0$$

El hecho de que $P(X = c) = 0$ cuando X es continua tiene una importante consecuencia práctica: la probabilidad de que X quede en algún intervalo entre a y b no depende de si el límite inferior a o el límite superior b está incluido en el cálculo de probabilidad:

$$P(a \le X \le b) = P(a < X < b) = P(a < X \le b) = P(a \le X < b) \tag{4.1}$$

Si X es discreta y tanto a como b son valores posibles (p. ej., X es binomial con $n = 20$ y $a = 5$, $b = 10$), entonces todas las cuatro probabilidades en (4.1) son diferentes.

La condición de probabilidad cero tiene un análogo físico. Considérese una barra circular sólida con área de sección transversal $= 1$ pulg^2. Coloque la barra a lo largo de un eje de medición y supóngase que la densidad de la barra en cualquier punto x está dada por el valor $f(x)$ de una función de densidad. Entonces si la barra se rebana en los puntos a y b y este segmento se retira, la cantidad de masa eliminada es $\int_a^b f(x)\,dx$; si la barra se rebana exactamente en el punto c, no se elimina masa. Se asigna masa a segmentos de intervalo de la barra pero no a puntos individuales.

Ejemplo 4.5 "Intervalo de tiempo" en el flujo de tránsito es el tiempo transcurrido entre el tiempo en que un carro termina de pasar por un punto fijo y el instante en que el siguiente carro comienza a pasar por ese punto. Sea $X =$ el intervalo de tiempo para dos carros consecutivos seleccionados al azar en una autopista durante un periodo de tráfico intenso. La siguiente función de densidad de probabilidad de X es en esencia el sugerido en "The Statistical Properties of Freeway Traffic" (*Transp. Res.* vol. 11: 221−228):

$$f(x) = \begin{cases} .15e^{-.15(x-.5)} & x \ge .5 \\ 0 & \text{de lo contrario} \end{cases}$$

La gráfica de $f(x)$ se da en la figura 4.4; no hay ninguna densidad asociada con intervalos de tiempo de menos de .5 y la densidad del intervalo de tiempo decrece con rapidez (exponencialmente rápido) a medida que x se incrementa a partir de .5. Claramente, $f(x) \ge 0$; para demostrar que $\int_{-\infty}^{\infty} f(x)\,dx = 1$, se utiliza el resultado obtenido con cálculo integral $\int_a^{\infty} e^{-kx}\,dx = (1/k)e^{-k \cdot a}$. Entonces

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{.5}^{\infty} .15e^{-.15(x-.5)}\,dx = .15e^{.075}\int_{.5}^{\infty} e^{-.15x}\,dx$$

$$= .15e^{.075} \cdot \frac{1}{.15} e^{-(.15)(.5)} = 1$$

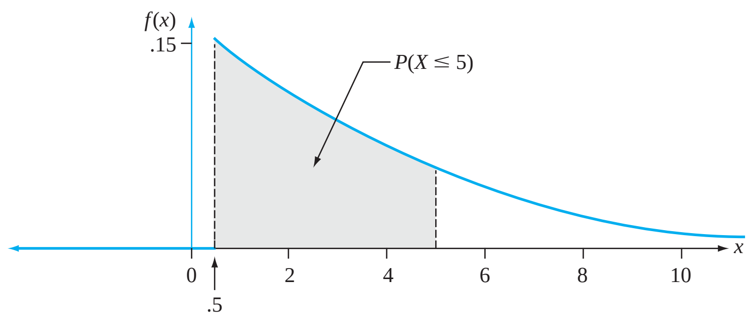


Figura 4.4 Curva de densidad del intervalo de tiempo entre vehículos en el ejemplo 4.5

La probabilidad de que el intervalo de tiempo sea cuando mucho de 5 segundos es

$$\begin{aligned} P(X \le 5) &= \int_{-\infty}^{5} f(x)\,dx = \int_{.5}^{5} .15e^{-.15(x-.5)}\,dx \\ &= .15e^{.075}\int_{.5}^{5} e^{-.15x}\,dx = .15e^{.075}\cdot\left(-\frac{1}{.15}e^{-.15x}\Big|_{x=.5}^{x=5}\right) \\ &= e^{.075}(-e^{-.75}+e^{-.075}) = 1.078(-.472+.928) = .491 \\ &= P(\text{menos de 5 s}) = P(X<5) \end{aligned}$$

A diferencia de las distribuciones discretas como la binomial, la hipergeométrica y la binomial negativa, la distribución de cualquier variable aleatoria continua dada no puede, en general, ser obtenida mediante argumentos probabilísticos. En cambio, se debe hacer una selección juiciosa de la función de densidad de probabilidad basada en conocimientos previos y en los datos disponibles. Afortunadamente, existen algunas familias generales de funciones de densidad de probabilidad que se ajustan bien a una amplia variedad de situaciones experimentales; varias de éstas se discuten más adelante en el capítulo.

Exactamente como en el caso discreto, a menudo es útil pensar en la población de interés como compuesta de valores X en lugar de individuos u objetos. La función de densidad de probabilidad es entonces un modelo de la distribución de valores en esta población numérica y con base en este modelo se pueden calcular varias características de la población (tal como la media).

EJERCICIOS Sección 4.1 (1–10)

1. La corriente en un circuito determinado, medido por un amperímetro es una variable aleatoria continua X con la función de densidad siguiente:

$$f(x) = \begin{cases} .075x + .2 & 3 \le x \le 5 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Grafique la función de densidad de probabilidad para verificar que el área total bajo la curva de densidad es de hecho 1.
b. Calcule $P(X \le 4)$. ¿Cómo se compara esta probabilidad con $P(X < 4)$?
c. Calcule $P(3.5 \le X \le 4.5)$ y $P(4.5 < X)$

2. Suponga que la temperatura de reacción X (en °C) en cierto proceso químico tiene una distribución uniforme con $A = -5$ y $B = 5$.
a. Calcule $P(X < 0)$.
b. Calcule $P(-2.5 < X < 2.5)$.
c. Calcule $P(-2 \le X \le 3)$.
d. Para que k satisfaga $-5 < k < k + 4 < 5$, calcule $P(k < X < k + 4)$.

3. El error implicado al hacer una medición es una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} .09375(4 - x^2) & -2 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la gráfica de $f(x)$.
b. Calcule $P(X > 0)$.
c. Calcule $P(-1 < X < 1)$.
d. Calcule $P(X < -.5 \text{ o } X > .5)$.

4. Sea X el esfuerzo vibratorio (lb/pulg^2) en el aspa de una turbina de viento a una velocidad del viento particular en un túnel aerodinámico. El artículo "Blade Fatigue Life Assessment with Application to VAWTS" (*J. of Solar Energy Engr.*, 1982: 107–111) propone la distribución de Rayleigh, con función de densidad de probabilidad

$$f(x;\theta) = \begin{cases} \dfrac{x}{\theta^2}\cdot e^{-x^2/(2\theta^2)} & x > 0 \\ 0 & \text{de lo contrario} \end{cases}$$

como modelo de la distribución X.
a. Verifique que $f(x; \theta)$ es una función de densidad de probabilidad legítima.
b. Suponga que $\theta = 100$ (un valor sugerido por una gráfica en el artículo). ¿Cuál es la probabilidad de que X sea cuando mucho de 200? ¿Menos de 200? ¿Por lo menos de 200?
c. ¿Cuál es la probabilidad de que X esté entre 100 y 200 (de nuevo suponiendo $\theta = 100$)?
d. Dé una expresión para $P(X \le x)$.

5. Un profesor universitario nunca termina su disertación antes del final de la hora y siempre termina dentro de dos minutos después de la hora. Sea X = el tiempo que transcurre entre el final de la hora y el final de la disertación y suponga que la función de densidad de probabilidad de X es

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Determine el valor de k y trace la curva de densidad correspondiente. [*Sugerencia*: el área total bajo la gráfica de $f(x)$ es 1.]
b. ¿Cuál es la probabilidad de que la disertación termine dentro de 1 minuto del final de la hora?

c. ¿Cuál es la probabilidad de que la disertación continúe después de la hora durante entre 60 y 90 segundos?
d. ¿Cuál es la probabilidad de que la disertación continúe durante por lo menos 90 segundos después del final de la hora?

6. El peso real de lectura de una pastilla de estéreo ajustado a 3 gramos en un tocadiscos particular puede ser considerado como una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} k[1 - (x - 3)^2] & 2 \le x \le 4 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la gráfica de $f(x)$.
b. Determine el valor de k.
c. ¿Cuál es la probabilidad de que el peso real de lectura sea mayor que el peso prescrito?
d. ¿Cuál es la probabilidad de que el peso real de lectura esté dentro de .25 gramo del peso prescrito?
e. ¿Cuál es la probabilidad de que el peso real difiera del peso prescrito por más de .5 gramo?

7. Se cree que el tiempo X (minutos) para que un ayudante de laboratorio prepare el equipo para cierto experimento tiene una distribución uniforme con $A = 25$ y $B = 35$.
a. Determine la función de densidad de probabilidad de X y trace la curva de densidad correspondiente.
b. ¿Cuál es la probabilidad de que el tiempo de preparación exceda de 33 minutos?
c. ¿Cuál es la probabilidad de que el tiempo de preparación esté dentro de 2 min del tiempo medio? [*Sugerencia*: identifique μ en la gráfica de $f(x)$.]
d. Para cualquier a tal que $25 < a < a + 2 < 35$, ¿cuál es la probabilidad de que el tiempo de preparación sea de entre a y $a + 2$ minutos?

8. Para ir al trabajo, un profesor primero debe tomar un camión cerca de su casa y luego tomar un segundo camión. Si el tiempo de espera (en minutos) en cada parada tiene una distribución uniforme con $A = 0$ y $B = 5$, entonces se puede demostrar que el tiempo de espera total Y tiene la función de densidad de probabilidad

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \le y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \le y \le 10 \\ 0 & y < 0 \text{ o } y > 10 \end{cases}$$

a. Trace la gráfica de la función de densidad de probabilidad de Y.
b. Verifique que $\int_{-\infty}^{\infty} f(y)\,dy = 1$.
c. ¿Cuál es la probabilidad de que el tiempo de espera total sea cuando mucho de 3 min?
d. ¿Cuál es la probabilidad de que el tiempo de espera total sea cuando mucho de 8 min?
e. ¿Cuál es la probabilidad de que el tiempo de espera total sea de entre 3 y 8 min?
f. ¿Cuál es la probabilidad de que el tiempo de espera total sea de menos de 2 min o de más de 6 min?

9. Considere de nuevo la función de densidad de probabilidad de X = intervalo de tiempo dado en el ejemplo 4.5. ¿Cuál es la probabilidad de que el intervalo de tiempo sea
a. cuando mucho de 6 segundos?
b. de más de 6 segundos? ¿Por lo menos de 6 segundos?
c. de entre 5 y 6 segundos?

10. Una familia de funciones de densidad de probabilidad que ha sido utilizada para aproximar la distribución del ingreso, el tamaño de la población de una ciudad y el tamaño de compañías es la familia Pareto. La familia tiene dos parámetros, k y θ, ambos > 0, y la función de densidad de probabilidad es

$$f(x; k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & x \ge \theta \\ 0 & x < \theta \end{cases}$$

a. Trace la gráfica de $f(x; k, \theta)$.
b. Verifique que el área total bajo la gráfica es igual a 1.
c. Si la variable aleatoria X tiene una función de densidad de probabilidad $f(x; k, \theta)$, con cualquier $b > \theta$ fija, obtenga una expresión para $P(X \le b)$.
d. Para $\theta < a < b$, obtenga una expresión para la probabilidad $P(a \le X \le b)$.

4.2 Funciones de distribución acumulativa y valores esperados

Varios de los más importantes conceptos introducidos en el estudio de distribuciones discretas también desempeñan un importante papel en las distribuciones continuas. Definiciones análogas a las del capítulo 3 implican reemplazar la suma por integración.

Función de distribución acumulativa

La función de distribución acumulativa $F(x)$ de una variable aleatoria discreta X da, con cualquier número especificado x, la probabilidad $P(X \le x)$. Se obtiene sumando la función de masa de probabilidad $p(y)$ a lo largo de todos los valores posibles y que satisfagan $y \le x$. La función de distribución acumulativa de una variable aleatoria continua da las mismas probabilidades $P(X \le x)$ y se obtiene integrando la función de densidad de probabilidad $f(y)$ entre los límites $-\infty$ y x.

DEFINICIÓN

La **función de distribución acumulativa** $F(x)$ de una variable aleatoria continua X se define para todo número x como

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$$

Para cada x, $F(x)$ es el área bajo la curva de densidad a la izquierda de x. Esto se ilustra en la figura 4.5, donde $F(x)$ se incrementa con suavidad a medida que x se incrementa.

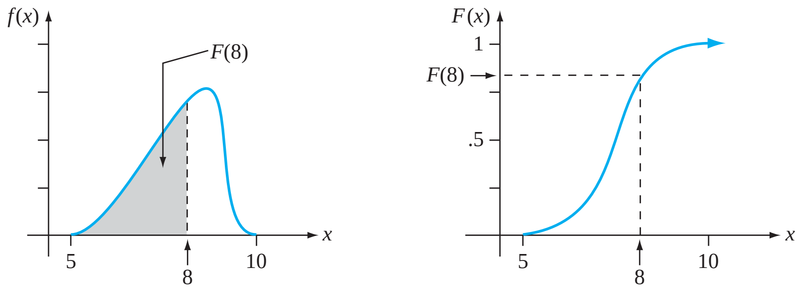


Figura 4.5 Una función de densidad de probabilidad y función de distribución acumulativa asociada

Ejemplo 4.6 Sea X el espesor de una cierta lámina de metal con distribución uniforme en $[A, B]$. La función de densidad se muestra en la figura 4.6. Para $x < A$, $F(x) = 0$, dado que no hay área bajo la gráfica de la función de densidad a la izquierda de la x. Con $x \geq B$, $F(x) = 1$, puesto que toda el área está acumulada a la izquierda de la x. Finalmente para $A \leq x \leq B$,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A}\,dy = \frac{1}{B-A} \cdot y \Big|_{y=A}^{y=x} = \frac{x-A}{B-A}$$

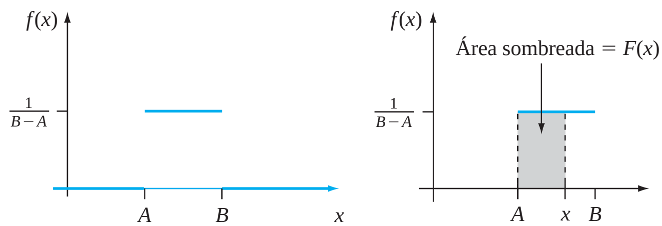


Figura 4.6 Función de densidad de probabilidad de una distribución uniforme

La función de distribución acumulativa completa es

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

La gráfica de esta función de distribución acumulativa aparece en la figura 4.7.

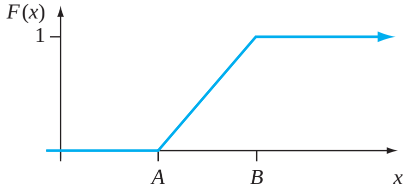


Figura 4.7 Función de distribución acumulativa de una distribución uniforme ■

Utilización de $F(x)$ para calcular probabilidades

La importancia de la función de distribución acumulativa en este caso, lo mismo que para variables aleatorias discretas, es que las probabilidades de varios intervalos pueden ser calculadas con una fórmula o tabla de $F(x)$.

PROPOSICIÓN

> Sea X una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y función de distribución acumulativa $F(x)$. Entonces para cualquier número a,
> $$P(X > a) = 1 - F(a)$$
> y para dos números cualesquiera a y b con $a < b$,
> $$P(a \leq X \leq b) = F(b) - F(a)$$

La figura 4.8 ilustra la segunda parte de esta proposición; la probabilidad deseada es el área sombreada bajo la curva de densidad entre a y b y es igual a la diferencia entre las dos áreas acumulativas sombreadas. Esto es diferente de lo que es apropiado para una variable aleatoria discreta de valor entero (p. ej., binomial o Poisson): $P(a \leq X \leq b) = F(b) - F(a - 1)$ cuando a y b son enteros.

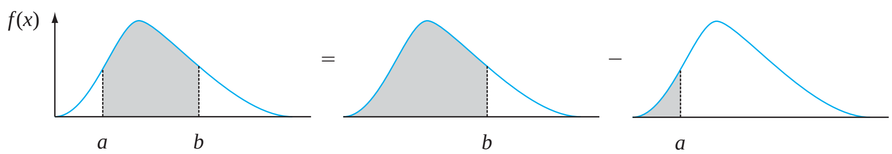


Figura 4.8 Cálculo de $P(a \leq X \leq b)$ a partir de probabilidades acumulativas

Ejemplo 4.7 Suponga que la función de densidad de probabilidad de la magnitud X de una carga dinámica sobre un puente (en newtons) está dada por

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{de lo contrario} \end{cases}$$

Para cualquier número x entre 0 y 2,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x}\left(\frac{1}{8} + \frac{3}{8}y\right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

Por lo tanto

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & 0 \leq x \leq 2 \\ 1 & 2 < x \end{cases}$$

Las gráficas de $f(x)$ y $F(x)$ se muestran en la figura 4.9. La probabilidad de que la carga sea de entre 1 y 1.5 es

$$\begin{aligned} P(1 \leq X \leq 1.5) &= F(1.5) - F(1) \\ &= \left[\frac{1}{8}(1.5) + \frac{3}{16}(1.5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] \\ &= \frac{19}{64} = .297 \end{aligned}$$

La probabilidad de que la carga sea de más de 1 es

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - F(1) = 1 - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] \\ &= \frac{11}{16} = .688 \end{aligned}$$

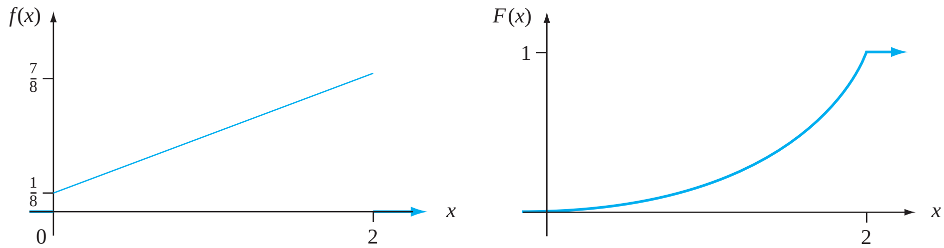


Figura 4.9 Función de densidad de probabilidad y función de distribución acumulativa del ejemplo 4.7 ■

Una vez que se obtiene la función de distribución acumulativa, cualquier probabilidad que implique X es fácil de calcular sin alguna integración adicional.

Obtención de $f(x)$ a partir de $F(x)$

Para X discreta, la función de masa de probabilidad se obtiene a partir de la función de distribución acumulativa considerando la diferencia entre dos valores $F(x)$. El análogo continuo de una diferencia es una derivada. El siguiente resultado es una consecuencia del teorema fundamental del cálculo.

PROPOSICIÓN

Si X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y función de distribución acumulativa $F(x)$, entonces en cada x que hace posible que la derivada $F'(x)$ exista, $F'(x) = f(x)$.

Ejemplo 4.8 (Continuación del ejemplo 4.6)

Cuando X tiene una distribución uniforme, $F(x)$ es diferenciable excepto en $x = A$ y $x = B$, donde la gráfica de $F(x)$ tiene esquinas afiladas. Como $F(x) = 0$ para $x < A$ y $F(x) = 1$ para $x > B$, $F'(x) = 0 = f(x)$ con dicha x. Para $A < x < B$,

$$F'(x) = \frac{d}{dx}\left(\frac{x - A}{B - A}\right) = \frac{1}{B - A} = f(x)$$ ■

Percentiles de una distribución continua

Cuando se dice que la calificación de un individuo en una prueba estaba en el 85º percentil de la población, significa que el 85% de todas las calificaciones de la población estuvieron por debajo de dicha calificación y que el 15% estuvo arriba. Asimismo, el 40º percentil es la calificación que sobrepasa al 40% de todas las calificaciones y es superado por el 60% de todas las calificaciones.

DEFINICIÓN

Sea p un número entre 0 y 1. El **(100*p*)° percentil** de la distribución de una variable aleatoria continua X, denotada por $\eta(p)$, se define como

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy \tag{4.2}$$

De acuerdo con la expresión (4.2), $\eta(p)$ es ese valor sobre el eje de medición de tal suerte que el $100p\%$ del área bajo la gráfica de $f(x)$ queda a la izquierda de $\eta(p)$ y $100(1 - p)\%$ queda a la derecha. Por lo tanto, $\eta(.75)$, el 75avo percentil, es tal que el área bajo la gráfica de $f(x)$ a la izquierda de $\eta(.75)$ es .75. La figura 4.10 ilustra la definición.

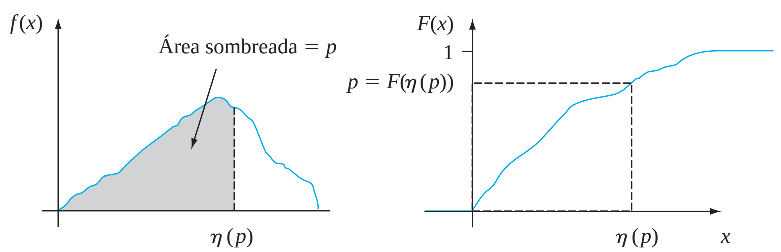


Figura 4.10 El (100*p*)° percentil de una distribución continua

Ejemplo 4.9 La distribución de la cantidad de grava (en toneladas) vendida por una compañía de materiales para la construcción particular en una semana dada es una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

La función de distribución acumulativa de las ventas para cualquier x entre 0 y 1 es

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)\,dy = \frac{3}{2}\left(y - \frac{y^3}{3}\right)\Bigg|_{y=0}^{y=x} = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

Las gráficas tanto de $f(x)$ como de $F(x)$ aparecen en la figura 4.11. El $(100p)^{\text{o}}$ percentil de esta distribución satisface la ecuación

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

es decir,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

Para el 50º percentil, $p = .5$ y la ecuación que se tiene que resolver es $\eta^3 - 3\eta + 1 = 0$; la solución es $\eta = \eta(.5) = .347$. Si la distribución no cambia de una semana a otra, entonces a la larga 50% de todas las semanas se realizarán ventas de menos de .347 ton y 50% de más de .347 ton.

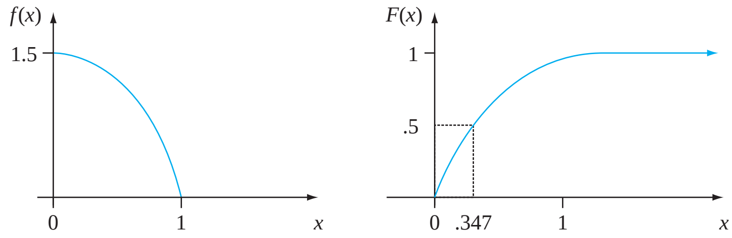


Figura 4.11 Función de densidad de probabilidad y función de distribución acumulativa del ejemplo 4.9 ■

DEFINICIÓN

La **mediana** de una distribución continua, denotada por $\tilde{\mu}$, es el 50º percentil, así que $\tilde{\mu}$ satisface $.5 = F(\tilde{\mu})$. Es decir, la mitad del área bajo la curva de densidad se encuentra a la izquierda de $\tilde{\mu}$ y la mitad a la derecha de $\tilde{\mu}$.

Una distribución continua cuya función de densidad de probabilidad es **simétrica** —la gráfica de la función de densidad de probabilidad a la izquierda de algún punto es una imagen de espejo de la gráfica a la derecha de dicho punto—, tiene una mediana $\tilde{\mu}$ igual al punto de simetría, puesto que la mitad del área bajo la curva queda a uno u otro lado de este punto. La figura 4.12 da varios ejemplos. A menudo se supone que el error en la medición de una cantidad física tiene una distribución simétrica.

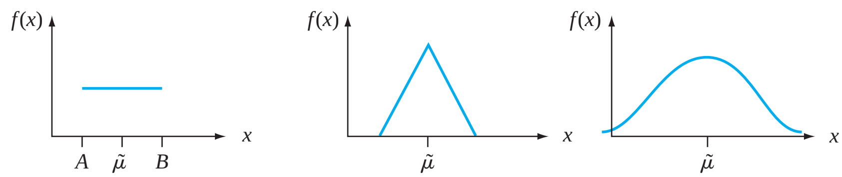


Figura 4.12 Medianas de distribuciones simétricas

Valores esperados

Para una variable aleatoria discreta X, $E(X)$ se obtuvo sumando $x \cdot p(x)$ a lo largo de posibles valores de X. Aquí se reemplaza la suma por la integración y la función de masa de probabilidad por la función de densidad de probabilidad para obtener un promedio ponderado continuo.

DEFINICIÓN

El **valor esperado** o **valor medio** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ es

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Ejemplo 4.10 (Continuación del ejemplo 4.9)

La función de densidad de probabilidad de las ventas semanales de grava X fue

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

por tanto

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)\,dx$$

$$= \frac{3}{2}\int_0^1 (x - x^3)\,dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{3}{8} \qquad \blacksquare$$

Cuando la función de densidad de probabilidad $f(x)$ especifica un modelo para la distribución de valores en una población numérica, entonces μ es la media de la población, la cual es la medida más frecuentemente utilizada de la ubicación o centro de la población.

Con frecuencia se desea calcular el valor esperado de alguna función $h(X)$ de la variable aleatoria X. Si se piensa en $h(X)$ como una nueva variable aleatoria Y, se utilizan técnicas de estadística matemática para obtener la función de densidad de probabilidad de Y, y $E(Y)$ se calcula a partir de la definición. Afortunadamente, como en el caso discreto, existe una forma más fácil de calcular $E[h(X)]$.

PROPOSICIÓN

Si X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y $h(X)$ es cualquier función de X, entonces

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\,dx$$

Ejemplo 4.11 Dos especies compiten en una región por el control de una cantidad limitada de cierto recurso. Sea X = la proporción del recurso controlado por la especie 1 y suponga que la función de densidad de probabilidad de X es

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

la cual es una distribución uniforme en [0, 1]. (En su libro *Ecological Diversity*, E. C. Pielou llama a esto el modelo del "palo roto" para la asignación de recursos, puesto que es análogo a la ruptura de un palo en un lugar seleccionado al azar.) Entonces la especie que controla la mayor parte de este recurso controla la cantidad

$$h(X) = \text{máx}\,(X, 1 - X) = \begin{cases} 1 - X & \text{si } 0 \le X < \dfrac{1}{2} \\ X & \text{si } \dfrac{1}{2} \le X \le 1 \end{cases}$$

La cantidad esperada controlada por la especie que controla la mayor parte es entonces

$$E[h(X)] = \int_{-\infty}^{\infty} \text{máx}(x, 1 - x) \cdot f(x)\,dx = \int_0^1 \text{máx}(x, 1 - x) \cdot 1\,dx$$

$$= \int_0^{1/2} (1 - x) \cdot 1\,dx + \int_{1/2}^1 x \cdot 1\,dx = \frac{3}{4} \qquad \blacksquare$$

Para $h(X)$, una función lineal, $E[h(X)] = E(aX + b) = aE(X) + b$.

En el caso discreto, la varianza de X se definió como la desviación al cuadrado esperada con respecto a μ y se calculó por medio de suma. En este caso de nuevo la integración reemplaza a la suma.

DEFINICIÓN

La **varianza** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ y valor medio μ es

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es $\sigma_X = \sqrt{V(X)}$.

La varianza y la desviación estándar dan medidas cuantitativas de cuánta dispersión hay en la distribución o población de valores x. Una vez más σ es aproximadamente del tamaño de una desviación típica de μ. El cálculo de σ^2 se facilita mediante el uso de la fórmula abreviada similar a la utilizada en el caso discreto.

PROPOSICIÓN

$$V(X) = E(X^2) - [E(X)]^2$$

Ejemplo 4.12 (Continuación del ejemplo 4.10)

Para X = ventas semanales de grava, se calculó $E(X) = \frac{3}{8}$. Como

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)\, dx$$

$$= \int_0^1 \frac{3}{2}(x^2 - x^4)\, dx = \frac{1}{5}$$

$$V(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} = .059 \text{ y } \sigma_X = .244$$ ■

Cuando $h(X) = aX + b$, el valor esperado y la varianza de $h(X)$ satisfacen las mismas propiedades que en el caso discreto: $E[h(X)] = a\mu + b$ y $V[h(X)] = a^2 \cdot \sigma^2$.

EJERCICIOS Sección 4.2 (11–27)

11. Sea X la cantidad de tiempo que un libro en reserva de dos horas está realmente prestado y supongamos que la función de distribución acumulativa es

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Use la función de distribución acumulativa para calcular lo siguiente:

a. $P(X \le 1)$
b. $P(.5 \le X \le 1)$
c. $P(X > 1.5)$
d. La mediana del tiempo de préstamo $\tilde{\mu}$ [resolver $.5 = F(\tilde{\mu})$]
e. $F'(x)$ para obtener la función de densidad $f(x)$
f. $E(X)$
g. $V(X)$ y σ_X
h. Si al prestatario se le cobra una cantidad $h(X = X^2$ cuando el tiempo de préstamo es X, calcule el cobro esperado $E[h(X)]$.

12. La función de distribución acumulativa de X (= error de medición) del ejercicio 3 es

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right) & -2 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

a. Calcule $P(X < 0)$.
b. Calcule $P(-1 < X < 1)$.
c. Calcule $P(.5 < X)$.
d. Verifique que $f(x)$ es la dada en el ejercicio 3 obteniendo $F'(x)$.
e. Verifique que $\tilde{\mu} = 0$.

13. El ejemplo 4.5 introdujo el concepto de intervalo de tiempo en el flujo de tránsito y propuso una distribución particular para X = el intervalo de tiempo entre dos carros consecutivos seleccionados al azar (s). Suponga que en un entorno de tránsito diferente, la distribución del intervalo de tiempo tiene la forma

$$f(x) = \begin{cases} \dfrac{k}{x^4} & x > 1 \\ 0 & x \le 1 \end{cases}$$

a. Determine el valor de k con el cual $f(x)$ es una función de densidad de probabilidad legítima.
b. Obtenga la función de distribución acumulativa.
c. Use la función de distribución acumulativa de (b) para determinar la probabilidad de que el intervalo de tiempo exceda de 2 segundos y también la probabilidad de que el intervalo sea de entre 2 y 3 segundos.
d. Obtenga un valor medio del intervalo de tiempo y su desviación estándar.
e. ¿Cuál es la probabilidad de que el intervalo de tiempo quede dentro de 1 desviación estándar del valor medio?

14. El artículo "Modeling Sediment and Water Column Interactions for Hidrophobic Pollutants" (*Water Research*, 1984: 1169–1174) sugiere la distribución uniforme en el intervalo (7.5, 20) como modelo de profundidad (cm) de la capa de bioturbación en sedimento en una región.
a. ¿Cuáles son la media y la varianza de la profundidad?
b. ¿Cuál es la función de distribución acumulativa de la profundidad?
c. ¿Cuál es la probabilidad de que la profundidad observada sea cuando mucho de 10? ¿Entre 10 y 15?
d. ¿Cuál es la probabilidad de que la profundidad observada esté dentro de 1 desviación estándar del valor medio? ¿Dentro de 2 desviaciones estándar?

15. Sea X la cantidad de espacio ocupado por un artículo colocado en un contenedor de 1 pie^3. La función de densidad de probabilidad de X es

$$f(x) = \begin{cases} 90x^8(1 - x) & 0 < x < 1 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Grafique la función de densidad de probabilidad. Luego obtenga la función de distribución acumulativa de X y grafíquela.
b. ¿Cuál es $P(X \le .5)$ [es decir, $F(.5)$]?
c. Con la función de distribución acumulativa de (a), ¿cuál es $P(.25 < X \le .5)$? ¿Cuál es $P(.25 \le X \le .5)$?
d. ¿Cuál es el 75avo percentil de la distribución?
e. Calcule $E(X)$ y σ_X.
f. ¿Cuál es la probabilidad de que X esté a más de 1 desviación estándar de su valor medio?

16. Responda los incisos (a)–(f) del ejercicio 15 con X = tiempo de disertación después de la hora dado en el ejercicio 5.

17. Si la distribución de X en el intervalo $[A, B]$ es uniforme
a. Obtenga una expresión para el $(100p)$º percentil.
b. Calcule $E(X)$, $V(X)$ y σ_X.
c. Con n, un entero positivo, calcule $E(X^n)$.

18. Sea X el voltaje a la salida de un micrófono y suponga que X tiene una distribución uniforme en el intervalo de -1 a 1. El voltaje es procesado por un "limitador duro" con valores de corte de $-.5$ y .5, de modo que la salida del limitador es una variable aleatoria Y relacionada con X por $Y = X$ si $|X| \le .5$, $Y = .5$ si $X > .5$ y $Y = -.5$ si $X < -.5$.
a. ¿Cuál es $P(Y = .5)$?
b. Obtenga la función de distribución acumulativa de Y y grafíquela.

19. Sea X una variable aleatoria continua con función de distribución acumulativa

$$F(x) = \begin{cases} 0 & x \le 0 \\ \dfrac{x}{4}\left[1 + \ln\left(\dfrac{4}{x}\right)\right] & 0 < x \le 4 \\ 1 & x > 4 \end{cases}$$

[Este tipo de función de distribución acumulativa es sugerido en el artículo "Variability in Measured Bedload-Transport Rates" (*Water Resources Bull.*, 1985: 39–48) como modelo de cierta variable hidrológica.] ¿Cuál es
a. $P(X \le 1)$?
b. $P(1 \le X \le 3)$?
c. La función de densidad de probabilidad de X?

20. Considere la función de densidad de probabilidad del tiempo de espera total Y de dos camiones

$$f(y) = \begin{cases} \dfrac{1}{25}y & 0 \le y < 5 \\ \dfrac{2}{5} - \dfrac{1}{25}y & 5 \le y \le 10 \\ 0 & \text{de lo contrario} \end{cases}$$

introducida en el ejercicio 8.
a. Calcule y trace la función de distribución acumulativa de Y. [*Sugerencia*: considere por separado $0 \le y < 5$ y $5 \le y \le 10$ al calcular $F(y)$. Una gráfica de la función de densidad de probabilidad debe ser útil.]
b. Obtenga una expresión para el $(100p)$º percentil. [*Sugerencia*: Considere por separado $0 < p < .5$ y $.5 < p < 1$.]
c. Calcule $E(Y)$ y $V(Y)$. ¿Cómo se comparan estos valores con el tiempo de espera probable y la varianza de un solo camión cuando el tiempo está uniformemente distribuido en [0, 5]?

21. Un ecólogo desea marcar una región de muestreo circular de 10 m de radio. Sin embargo, el radio de la región resultante en realidad es una variable aleatoria R con función de densidad de probabilidad

$$f(r) = \begin{cases} \dfrac{3}{4}[1 - (10 - r)^2] & 9 \le r \le 11 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es el área esperada de la región circular resultante?

22. La demanda semanal de gas propano (en miles de galones) de una instalación particular es una variable aleatoria X con función de densidad de probabilidad

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Calcule la función de distribución acumulativa de X.
b. Obtenga una expresión para el $(100p)$º percentil. ¿Cuál es el valor de $\tilde{\mu}$?
c. Calcule $E(X)$ y $V(X)$.
d. Si 1500 galones están en existencia al principio de la semana y no se espera ningún nuevo suministro durante la semana, ¿cuántos de los 1500 galones se espera que queden al final de la semana? [*Sugerencia*: sea $h(x)$ = cantidad que queda cuando la demanda = x.]

23. Si la temperatura a la cual cierto compuesto se funde es una variable aleatoria con valor medio de 120°C y desviación estándar de 2°C, ¿cuáles son la temperatura media y la desviación estándar medidas en °F? [*Sugerencia*: °F = 1.8°C + 32.]

24. La función de densidad de probabilidad de Pareto de X es

$$f(x; k, \theta) = \begin{cases} \dfrac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

introducida en el ejercicio 10.

a. Si $k > 1$, calcule $E(X)$.
b. ¿Qué se puede decir sobre $E(X)$ si $k = 1$?
c. Si $k > 2$, demuestre que $V(X) = k\theta^2(k - 1)^{-2}(k - 2)^{-1}$.
d. Si $k = 2$, ¿qué se puede decir sobre $V(X)$?
e. ¿Qué condiciones en cuanto a k son necesarias para garantizar que $E(X^n)$ es finito?

25. Sea X la temperatura en °C a la cual ocurre una reacción química y sea Y la temperatura en °F (así que $Y = 1.8X + 32$).
a. Si la mediana de la distribución X es $\tilde{\mu}$, demuestre que $1.8\tilde{\mu} + 32$ es la mediana de la distribución Y.
b. ¿Cómo está relacionado el 90º percentil de la distribución Y con el 90º percentil de la distribución X? Verifique su conjetura.
c. Más generalmente, si $Y = aX + b$, ¿cómo está relacionado cualquier percentil de la distribución Y con el percentil correspondiente de la distribución X?

26. Sea X los gastos médicos totales (en miles de dólares) en que incurre un individuo particular durante un año dado. Aunque X es una variable aleatoria discreta, suponga que su distribución es bastante bien aproximada por una distribución continua con función de densidad de probabilidad $f(x) = k(1 + x/2.5)^{-7}$ para $x \geq 0$.
a. ¿Cuál es el valor de k?
b. Grafique la función de densidad de probabilidad de X.
c. ¿Cuáles son el valor esperado y la desviación estándar de los gastos médicos totales?
d. Este individuo está cubierto por un plan de aseguramiento que le impone una provisión deducible de \$500 (así que los primeros \$500 de gastos son pagados por el individuo). Luego el plan pagará 80% de cualquier gasto adicional que exceda de \$500 y el pago máximo por parte del individuo (incluida la cantidad deducible) es de \$2500. Sea Y la cantidad de gastos médicos de este individuo pagados por la compañía de seguros. ¿Cuál es el valor esperado de Y? [*Sugerencia*: primero indague qué valor de X corresponde al gasto máximo que sale del bolsillo de \$2500. Luego escriba una expresión para Y como una función de X (la cual implique varios precios diferentes) y calcule el valor esperado de la función.]

27. Cuando se lanza un dardo a un blanco circular, considere la ubicación del punto de aterrizaje con respecto al centro del blanco. Sea X el ángulo en grados medido con respecto a la horizontal y suponga que X está uniformemente distribuida en [0, 360]. Defina Y como la variable transformada $Y = h(X) = (2\pi/360)X - \pi$, por lo tanto, Y es el ángulo medido en radianes y Y está entre $-\pi$ y π. Obtenga $E(Y)$ y σ_Y obteniendo primero $E(X)$ y σ_X y luego utilizando el hecho de que $h(X)$ es una función lineal de X.

4.3 Distribución normal

La **distribución normal** es la más importante en toda la probabilidad y estadística. Muchas poblaciones numéricas tienen distribuciones que pueden ser representadas muy fielmente por una curva normal apropiada. Los ejemplos incluyen estaturas, pesos y otras características físicas (el famoso artículo *Biometrika* 1903 "On the Laws of Inheritance in Man" discutió muchos ejemplos de esta clase), errores de medición en experimentos científicos, mediciones antropométricas en fósiles, tiempos de reacción en experimentos psicológicos, mediciones de inteligencia y aptitud, calificaciones en varios exámenes y numerosas medidas e indicadores económicos. Además, aun cuando las variables individuales no estén normalmente distribuidas, las sumas y promedios de las variables en condiciones adecuadas tendrán de manera aproximada una distribución normal; éste es el contenido del teorema del límite central discutido en el siguiente capítulo.

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución normal** con parámetros μ y σ (o μ y σ^2), donde $-\infty < \mu < \infty$ y $0 < \sigma$, si la función de densidad de probabilidad de X es

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{4.3}$$

De nuevo e denota la base del sistema de logaritmos naturales y es aproximadamente igual a 2.71828 y π representa la conocida constante matemática con un valor aproximado de 3.14159. El enunciado de que X está normalmente distribuida con los parámetros μ y σ^2 a menudo se abrevia como $X \sim N(\mu, \sigma^2)$.

Claramente $f(x; \mu, \sigma) \geq 0$ aunque se tiene que utilizar un argumento de cálculo un tanto complicado para verificar que $\int_{-\infty}^{\infty} f(x; \mu, \sigma)\, dx = 1$. Se puede demostrar que $E(X) = \mu$ y $V(X) = \sigma^2$, de modo que los parámetros son la media y la desviación estándar de X. La figura 4.13 representa gráficas de $f(x; \mu, \sigma)$ de varios pares diferentes (μ, σ). Cada curva de densidad es simétrica con respecto a μ y acampanada, de modo que el centro de la campana (punto de simetría) es tanto la media de la distribución como la mediana. El valor de σ es la distancia desde μ hasta los puntos de inflexión de la curva (los puntos donde la curva cambia de dirección hacia abajo o hacia arriba). Los grandes valores de σ producen gráficas que están bastante extendidas en torno a μ, en tanto que los valores pequeños de σ dan gráficas con una alta cresta sobre μ y la mayor parte del área bajo la gráfica bastante cerca de μ. Así pues, una σ grande implica que se puede observar muy bien un valor de X alejado de μ, en tanto que dicho valor es bastante improbable cuando σ es pequeña.

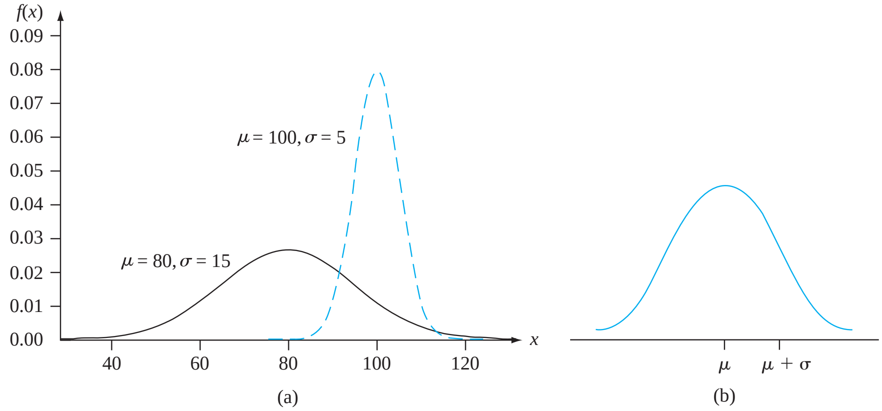


Figura 4.13 (a) Dos curvas diferentes de densidad normal (b) Visualización de μ y σ para una distribución normal

Distribución normal estándar

El cálculo de $P(a \leq X \leq b)$ cuando X es una variable aleatoria normal con parámetros μ y σ, requiere determinar

$$\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\, dx \tag{4.4}$$

Ninguna de las técnicas estándar de integración puede ser utilizada para lograr esto. En cambio, con $\mu = 0$ y $\sigma = 1$, se calculó la expresión (4.4) por medio de técnicas numéricas y se tabuló para ciertos valores de a y b. Esta tabla también puede ser utilizada para calcular probabilidades con cualesquiera otros valores de μ y σ considerados.

DEFINICIÓN

La distribución normal con valores de parámetro $\mu = 0$ y $\sigma = 1$ se llama **distribución normal estándar.** Una variable aleatoria que tiene una distribución normal estándar se llama **variable aleatoria normal estándar** y se denotará por Z. La función de densidad de probabilidad de Z es

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

La gráfica de $f(z; 0, 1)$ se llama curva *normal estándar* (o z). Sus puntos de inflexión estan en 1 y -1. La función de distribución acumulativa de Z es $P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$ la cual será denotada por $\Phi(z)$.

La distribución normal estándar no siempre sirve como modelo de una población que surge naturalmente. En cambio, es una distribución de referencia de la que se puede obtener información sobre otra distribución normal. La tabla A.3 del apéndice, da $\Phi(z) = P(Z \leq z)$, el área bajo la curva de densidad normal estándar a la izquierda de z con $z = -3.49, -3.48, \ldots, 3.48, 3.49$. La figura 4.14 ilustra el tipo de área acumulativa (probabilidad) tabulada en la tabla A.3. Con esta tabla, varias otras probabilidades que implican Z pueden ser calculadas.

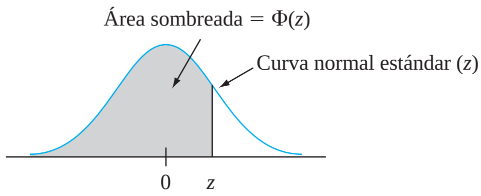


Figura 4.14 Áreas acumulativas normales estándar tabuladas en la tabla A.3 del apéndice

Ejemplo 4.13 Determínense las siguientes probabilidades normales estándar: (a) $P(Z \leq 1.25)$, (b) $P(Z > 1.25)$, (c) $P(Z \leq -1.25)$ y (d) $P(-.38 \leq Z \leq 1.25)$.

a. $P(Z \leq 1.25) = \Phi(1.25)$, una probabilidad tabulada en la tabla A.3 del apéndice en la intersección de la fila 1.2 y la columna .05. El número allí es .8944, así que $P(Z \leq 1.25) = .8944$. La figura 4.15(a) ilustra esta probabilidad.

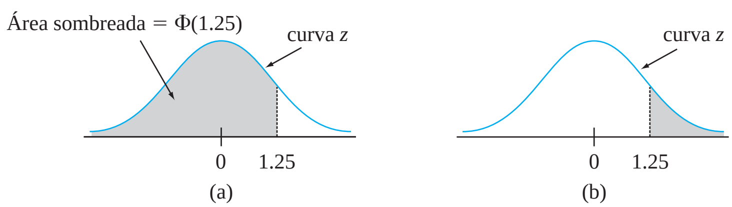


Figura 4.15 Áreas (probabilidades) de curvas normales del ejemplo 4.13

b. $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - \Phi(1.25)$, el área bajo la curva z a la derecha de 1.25 (un área de cola superior). En ese caso $\Phi(1.25) = .8944$ implica que $P(Z > 1.25) = .1056$. Como Z es una variable aleatoria continua, $P(Z \geq 1.25) = .1056$. Véase la figura 4.15(b).

c. $P(Z \leq -1.25) = \Phi(-1.25)$, un área de cola inferior. Directamente de la tabla A.3 del apéndice, $\Phi(-1.25) = .1056$. Por simetría de la curva z, ésta es la misma respuesta del inciso (b).

d. $P(-.38 \leq Z \leq 1.25)$ es el área bajo la curva normal estándar sobre el intervalo cuyo punto extremo izquierdo es $-.38$ y cuyo punto extremo derecho es 1.25. Según la sección 4.2, si X es una variable aleatoria continua con función de distribución acumulativa $F(x)$, entonces $P(a \leq X \leq b) = F(b) - F(a)$. Por lo tanto $P(-.38 \leq Z \leq 1.25) = \Phi(1.25) - \Phi(-.38) = .8944 - .3520 = .5424$. (Véase la figura 4.16.)

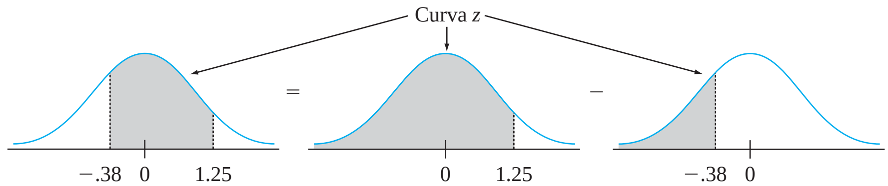


Figura 4.16 $P(-.38 \leq Z \leq 1.25)$ como la diferencia entre dos áreas acumulativas ■

Percentiles de la distribución normal estándar

Con cualquier p entre 0 y 1, se puede utilizar la tabla A.3 del apéndice para obtener el $(100p)^{o}$ percentil de la distribución normal estándar.

Ejemplo 4.14 El 99º percentil de la distribución normal estándar es ese valor sobre el eje horizontal tal que el área bajo la curva z a la izquierda de dicho valor es .9900. La tabla A.3 del apéndice da con z fija el área bajo la curva normal estándar a la izquierda de z, mientras que aquí se tiene el área y se desea el valor de z. Éste es el problema "inverso" a $P(Z \leq z) = ?$ así que la tabla se utiliza a la inversa: encuentre en la mitad de la tabla .9900; la fila y la columna en la que se encuentra identifica el 99° percentil z. En este caso .9901 queda en la intersección de la fila 2.3 y la columna .03, así que el 99° percentil es (aproximadamente) $z = 2.33$. (Véase la figura 4.17). Por simetría, el primer percentil está tan debajo de 0 como el 99° está sobre 0, así que es igual a -2.33 (1% queda debajo del primero y también sobre el 99°). (Véase la figura 4.18.)

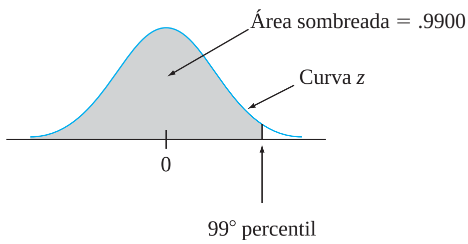


Figura 4.17 Localización del 99º percentil

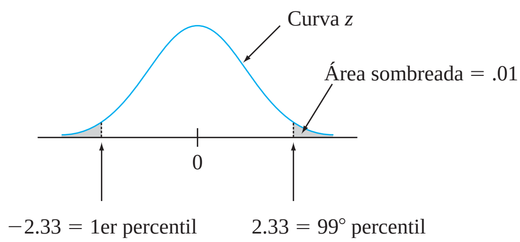


Figura 4.18 Relación entre el 1º y el 99º percentiles ■

En general, la fila y la columna de la tabla A.3 del apéndice, donde la entrada p está localizado identifican el $(100p)^{o}$ percentil (p. ej., el 67º percentil se obtiene localizando .6700 en el cuerpo de la tabla, la cual da $z = .44$). Si p no aparece, a menudo se utiliza el número más cercano a él, aunque la interpolación lineal da una respuesta más precisa. Por ejemplo, para encontrar el 95º percentil, se busca .9500 adentro de la tabla. Aunque .9500 no aparece, tanto .9495 como .9505 sí, correspondientes a $z = 1.64$ y 1.65, respectivamente.

Como .9500 está a la mitad entre las dos probabilidades que sí aparecen, se utilizará 1.645 como el 95º percentil y -1.645 como el 5º percentil.

Notación z_α para valores z críticos

En inferencia estadística se necesitan valores sobre el eje horizontal z que capturen ciertas áreas de cola pequeña bajo la curva normal estándar.

> **Notación**
>
> z_α denotará el valor sobre el eje z para el cual α del área bajo la curva z queda a la derecha de z_α. (Véase la figura 4.19.)

Por ejemplo, $z_{.10}$ captura el área de cola superior .10, y $z_{.01}$ captura el área de cola superior .01.

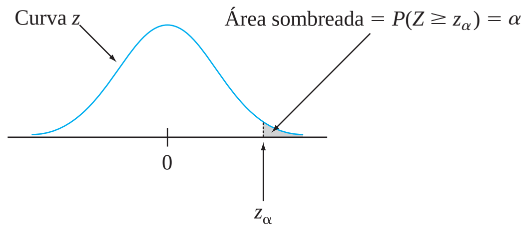


Figura 4.19 Notación z_α ilustrada

Como α del área bajo la curva z queda a la derecha de z_α, $1 - \alpha$ del área queda a su izquierda. Por lo tanto, z_α *es el* $100(1 - \alpha)$º *percentil de la distribución normal estándar*. Por simetría el área bajo la curva normal estándar a la izquierda de $-z_\alpha$ también es α. Los valores z_α en general se conocen como **valores críticos *z*.** La tabla 4.1 incluye los percentiles z y los valores z_α más útiles.

Tabla 4.1 Percentiles de la distribución normal estándar y valores críticos

Percentil	90	95	97.5	99	99.5	99.9	99.95
α (área de la cola)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1 - \alpha)$º percentil	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Ejemplo 4.15 $z_{.05}$ es el $100(1 - .05)$º $=$ 95º percentil de la distribución normal estándar, por lo tanto $z_{.05} = 1.645$. El área bajo la curva normal estándar a la izquierda de $-z_{.05}$ también es .05. (Véase la figura 4.20.)

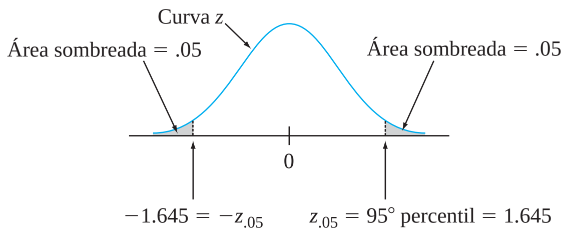


Figura 4.20 Determinación de $z_{.05}$ ■

Distribuciones normales no estándar

Cuando $X \sim N(\mu, \sigma^2)$, las probabilidades que implican X se calculan "estandarizando". La **variable estandarizada** es $(X - \mu)/\sigma$. Al restar μ la media cambia de μ a cero y luego al dividir entre σ cambian las escalas de la variable de modo que la desviación estándar es 1 en lugar de σ.

PROPOSICIÓN

Si X tiene una distribución normal con media μ y desviación estándar σ, entonces

$$Z = \frac{X - \mu}{\sigma}$$

tiene una distribución normal estándar. Por lo tanto

$$P(a \le X \le b) = P\left(\frac{a - \mu}{\sigma} \le Z \le \frac{b - \mu}{\sigma}\right)$$
$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \qquad P(X \ge b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

La idea clave de la proposición es que estandarizando, cualquier probabilidad que implique X puede ser expresada como una probabilidad que implica una variable aleatoria normal estándar Z, de modo que se pueda utilizar la tabla A.3 del apéndice. Esto se ilustra en la figura 4.21. La proposición se comprueba escribiendo la función de distribución acumulativa de $Z = (X - \mu)/\sigma$ como

$$P(Z \le z) = P(X \le \sigma z + \mu) = \int_{-\infty}^{\sigma z + \mu} f(x; \mu, \sigma)dx$$

Utilizando un resultado del cálculo, esta integral puede ser diferenciada con respecto a z para que dé la función de densidad de probabilidad deseada $f(z; 0, 1)$.

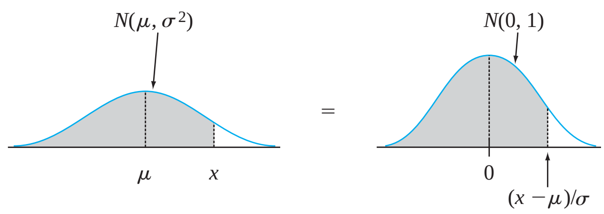


Figura 4.21 Igualdad de áreas de curvas normales estándar y no estándar

Ejemplo 4.16 El tiempo que requiere un conductor para reaccionar a las luces de freno de un vehículo que está desacelerando es crítico para evitar colisiones por alcance. El artículo "Fast-Rise Brake Lamp as a Collision-Prevention Device" (*Ergonomics*, 1993: 391–395), sugiere que el tiempo de reacción de respuesta en tráfico a una señal de luces de freno estándar puede ser modelado con una distribución normal que tiene un valor medio de 1.25 s y desviación estándar de .46 s. ¿Cuál es la probabilidad de que el tiempo de reacción sea de entre 1.00 y 1.75 segundos? Si X denota el tiempo de reacción, entonces estandarizando se obtiene

$$1.00 \le X \le 1.75$$

si y sólo si

$$\frac{1.00 - 1.25}{.46} \le \frac{X - 1.25}{.46} \le \frac{1.75 - 1.25}{.46}$$

Por lo tanto

$$\begin{aligned} P(1.00 \le X \le 1.75) &= P\left(\frac{1.00 - 1.25}{.46} \le Z \le \frac{1.75 - 1.25}{.46}\right) \\ &= P(-.54 \le Z \le 1.09) = \Phi(1.09) - \Phi(-.54) \\ &= .8621 - .2946 = .5675 \end{aligned}$$

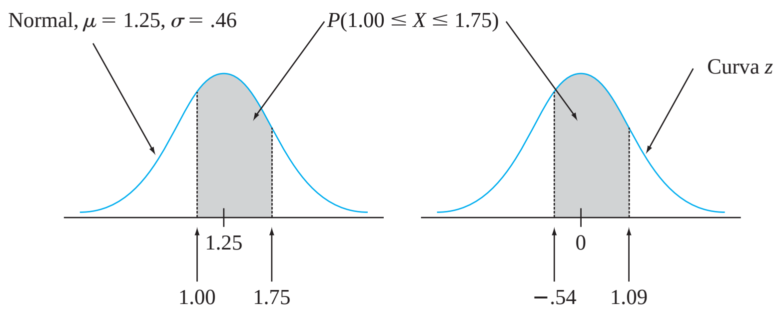


Figura 4.22 Curvas normales del ejemplo 4.16

Esto se ilustra en la figura 4.22. Asimismo, si se ven los 2 segundos como un tiempo de reacción críticamente largo, la probabilidad de que el tiempo de reacción real exceda este valor es

$$P(X > 2) = P\left(Z > \frac{2 - 1.25}{.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = .0516 \quad ■$$

Estandarizar cantidades no lleva a nada más que a calcular una distancia del valor medio y luego reexpresarla como algún número de desviaciones estándar. Por lo tanto, si $\mu = 100$ y $\sigma = 15$, entonces $x = 130$ corresponde a $z = (130 - 100)/15 = 30/15 = 2.00$. Es decir, 130 está a 2 desviaciones estándar sobre (a la derecha de) el valor medio. Asimismo, estandarizando 85 se obtiene $(85 - 100)/15 = -1.00$, por lo tanto 85 está a 1 desviación estándar por debajo de la media. La tabla z se aplica a *cualquier* distribución normal siempre que se piense en función del número de desviaciones estándar con respecto al valor medio.

Ejemplo 4.17 Se sabe que el voltaje de ruptura de un diodo seleccionado al azar de un tipo particular está normalmente distribuido. ¿Cuál es la probabilidad de que el voltaje de ruptura de un diodo esté dentro de 1 desviación estándar de su valor medio? Esta pregunta puede ser respondida sin conocer μ o σ, en tanto se sepa que la distribución es normal; la respuesta es la misma para *cualquier* distribución normal:

$$\begin{aligned} P(X \text{ está dentro de 1 desviación estándar de su media}) &= P(\mu - \sigma \le X \le \mu + \sigma) \\ &= P\left(\frac{\mu - \sigma - \mu}{\sigma} \le Z \le \frac{\mu + \sigma - \mu}{\sigma}\right) \\ &= P(-1.00 \le Z \le 1.00) \\ &= \Phi(1.00) - \Phi(-1.00) = .6826 \end{aligned}$$

La probabilidad de que X esté dentro de 2 desviaciones estándar de su media es $P(-2.00 \leq Z \leq 2.00) = .9544$ y dentro de 3 desviaciones estándar de su media es $P(-3.00 \leq Z \leq 3.00) = .9974$. ■

Los resultados del ejemplo 4.17 a menudo se reportan en forma de porcentaje y se les conoce como la *regla empírica* (porque la evidencia empírica ha demostrado que los histogramas de datos reales con frecuencia pueden ser aproximados por curvas normales).

Si la distribución de la población de una variable es (aproximadamente) normal, entonces

1. Aproximadamente 68% de los valores están dentro de 1 DE de la media.
2. Aproximadamente 95% de los valores están dentro de 2 DE de la media.
3. Aproximadamente 99.7% de los valores están dentro de 3 DE de la media.

En realidad es inusual observar un valor de una población normal que esté mucho más lejos de 2 desviaciones estándar de μ. Estos resultados serán importantes en el desarrollo de procedimientos de prueba de hipótesis en capítulos posteriores.

Percentiles de una distribución normal arbitraria

El $(100p)^{\text{o}}$ percentil de una distribución normal con media μ y desviación estándar σ es fácil de relacionar con el $(100p)^{\text{o}}$ percentil de la distribución normal estándar.

PROPOSICIÓN

$$\begin{matrix}(100p)^\circ \text{ percentil} \\ \text{para } (\mu, \sigma) \text{ normal}\end{matrix} = \mu + \left[\begin{matrix}(100p)^\circ \text{ para} \\ \text{normal estándar}\end{matrix}\right] \cdot \sigma$$

Otra forma de decir esto es que si z es el percentil deseado de la distribución normal estándar, entonces el percentil deseado de la distribución (μ, σ) normal está a z desviaciones estándar de μ.

Ejemplo 4.18 La cantidad de agua destilada despachada por cierta máquina está normalmente distribuida con valor medio de 64 oz y desviación estándar de .78 oz. ¿Qué tamaño de contenedor c asegurará que ocurra rebosamiento sólo .5% del tiempo? Si X denota la cantidad despachada, la condición deseada es que $P(X > c) = .005$ o, en forma equivalente, que $P(X \leq c) = .995$. Por lo tanto c es el 99.5º percentil de la distribución normal con $\mu = 64$ y $\sigma = .78$. El 99.5º percentil de la distribución normal estándar es 2.58, por lo tanto

$$c = \eta(.995) = 64 + (2.58)(.78) = 64 + 2.0 = 66 \text{ oz}$$

Esto se ilustra en la figura 4.23.

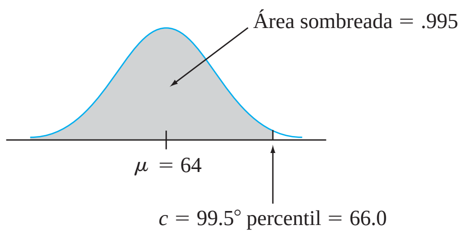


Figura 4.23 Distribución de la cantidad despachada en el ejemplo 4.18 ■

Distribución normal y poblaciones discretas

La distribución normal a menudo se utiliza como una aproximación a la distribución de valores en una población discreta. En semejantes situaciones se debe tener cuidado especial para asegurarse de que las probabilidades se calculen con precisión.

Ejemplo 4.19 Se sabe que el coeficiente intelectual en una población particular (medido con una prueba estándar) está más o menos normalmente distribuido con $\mu = 100$ y $\sigma = 15$. ¿Cuál es la probabilidad de que un individuo seleccionado al azar tenga un CI de por lo menos 125? Con $X =$ el CI de una persona seleccionada al azar, se desea $P(X \geq 125)$. La tentación en este caso es estandarizar $X \geq 125$ como en ejemplos previos. Sin embargo, la distribución de la población de coeficientes intelectuales en realidad es discreta, puesto que los coeficientes intelectuales son valores enteros. Así que la curva normal es una aproximación a un histograma de probabilidad discreto como se ilustra en la figura 4.24.

Los rectángulos del histograma están *centrados* en enteros, por lo que los coeficientes intelectuales de por lo menos 125 corresponden a rectángulos que comienzan en 124.5, la zona sombreada en la figura 4.24. Por lo tanto en realidad se desea el área bajo la curva aproximadamente normal a la derecha de 124.5. Si se estandariza este valor se obtiene $P(Z \geq 1.63) = .0516$, en tanto que si se estandariza 125 se obtiene $P(Z \geq 1.67) = .0475$. La diferencia no es grande, pero la respuesta .0516 es más precisa. Asimismo, $P(X = 125)$ sería aproximada por el área entre 124.5 y 125.5, puesto que el área bajo la curva normal sobre el valor único de 125 es cero.

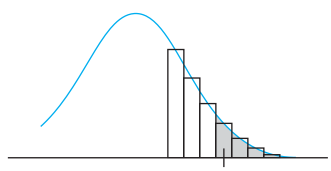


Figura 4.24 Aproximación normal a una distribución discreta ■

La corrección en cuanto a discrecionalidad de la distribución subyacente en el ejemplo 4.19 a menudo se llama **corrección de continuidad.** Es útil en la siguiente aplicación de la distribución normal al cálculo de probabilidades binomiales.

Aproximación de la distribución binomial

Recuérdese que el valor medio y la desviación estándar de una variable aleatoria binomial X son $\mu_X = np$ y $\sigma_X = \sqrt{npq}$, respectivamente. La figura 4.25 muestra un histograma de probabilidad binomial de la distribución binomial con $n = 20$, $p = .6$ con el cual $\mu = 20(.6) = 12$ y $\sigma = \sqrt{20(.6)(.4)} = 2.19$. Sobre el histograma de probabilidad se superpuso una curva normal con estas μ y σ. Aunque el histograma de probabilidad es un poco asimétrico (debido a que $p \neq .5$), la curva normal da una muy buena aproximación, sobre todo en la parte media de la figura. El área de cualquier rectángulo (probabilidad de cualquier valor X particular), excepto las de los localizados en las colas extremas, puede ser aproximada con precisión mediante el área de la curva normal correspondiente. Por ejemplo, $P(X = 10) = B(10; 20, .6) - B(9; 20, .6) = .117$, mientras que el área bajo la curva normal entre 9.5 y 10.5 es $P(-1.14 \leq Z \leq -.68) = .1212$.

Más generalmente, en tanto que el histograma de probabilidad binomial no sea demasiado asimétrico, las probabilidades binomiales pueden ser aproximadas muy bien por áreas de curva normal. Se acostumbra entonces decir que X tiene aproximadamente una distribución normal.

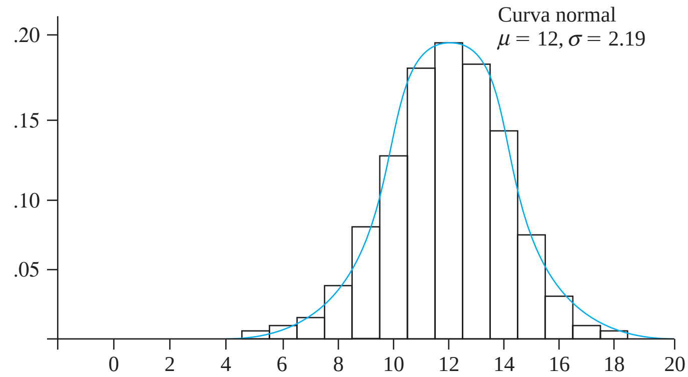


Figura 4.25 Histograma de probabilidad binomial para $n = 20$, $p = .6$ con curva de aproximación normal sobrepuesta

PROPOSICIÓN

Sea X una variable aleatoria binomial basada en n ensayos con probabilidad de éxito p. Entonces, si el histograma de probabilidad binomial no es demasiado asimétrico, X tiene aproximadamente una distribución normal con $\mu = np$ y $\sigma = \sqrt{npq}$. En particular, con $x =$ un valor posible de X,

$$P(X \leq x) = B(x, n, p) \approx \begin{pmatrix} \text{área bajo la curva normal} \\ \text{a la izquierda de } x + .5 \end{pmatrix}$$
$$= \Phi\left(\frac{x + .5 - np}{\sqrt{npq}}\right)$$

En la práctica, la aproximación es adecuada siempre que $np \geq 10$ y $nq \geq 10$, puesto que en ese caso existe bastante simetría en la distribución binomial subyacente.

Una comprobación directa de este resultado es bastante difícil. En el siguiente capítulo se verá que es una consecuencia de un resultado más general llamado teorema del límite central. Con toda honestidad, esta aproximación no es tan importante en el cálculo de probabilidad como una vez lo fue. Esto se debe a que los programas de computadora ahora son capaces de calcular probabilidades binomiales con exactitud para valores bastante grandes de n.

Ejemplo 4.20 Suponga que 25% de todos los estudiantes en una gran universidad pública reciben ayuda financiera. Sea X el número de estudiantes que reciben esta ayuda en una muestra aleatoria de tamaño 50, de modo que $p = .25$. Entonces $\mu = 12.5$ y $\sigma = 3.06$. Como $np = 50(.25) = 12.5 \geq 10$ y $nq = 37.5 \geq 10$, la aproximación puede ser aplicada con seguridad. La probabilidad de que a lo más 10 estudiantes reciban ayuda es

$$P(X \leq 10) = B(10; 50, .25) \approx \Phi\left(\frac{10 + .5 - 12.5}{3.06}\right)$$
$$= \Phi(-.65) = .2578$$

Asimismo, la probabilidad de que entre 5 y 15 (inclusive) de los estudiantes seleccionados reciban ayuda es

$$P(5 \leq X \leq 15) = B(15; 50, .25) - B(4; 50, .25)$$
$$\approx \Phi\left(\frac{15.5 - 12.5}{3.06}\right) - \Phi\left(\frac{4.5 - 12.5}{3.06}\right) = .8320$$

Las probabilidades exactas son .2622 y .8348, respectivamente, así que las aproximaciones son bastante buenas. En el último cálculo, la probabilidad $P(5 \leq X \leq 15)$ está siendo aproximada por el área bajo la curva normal entre 4.5 y 15.5; se utiliza la corrección de continuidad tanto para el límite superior como para el inferior. ■

Cuando el objetivo de la investigación es hacer una inferencia sobre una proporción de población p, el interés se enfocará en la proporción muestral de X/n éxitos y no en X. Como esta proporción es exactamente X multiplicada por la constante $1/n$, también tendrá aproximadamente una distribución normal (con media $\mu = p$ y desviación estándar $\sigma = \sqrt{pq/n}$, siempre que $np \geq 10$ y $nq \geq 10$. Esta aproximación normal es la base de varios procedimientos inferenciales que se discutirán en capítulos posteriores.

EJERCICIOS Sección 4.3 (28–58)

28. Sea Z una variable aleatoria normal estándar y calcule las siguientes probabilidades, trace las figuras siempre que sea apropiado.

a. $P(0 \leq Z \leq 2.17)$ **b.** $P(0 \leq Z \leq 1)$
c. $P(-2.50 \leq Z \leq 0)$ **d.** $P(-2.50 \leq Z \leq 2.50)$
e. $P(Z \leq 1.37)$ **f.** $P(-1.75 \leq Z)$
g. $P(-1.50 \leq Z \leq 2.00)$ **h.** $P(1.37 \leq Z \leq 2.50)$
i. $P(1.50 \leq Z)$ **j.** $P(|Z| \leq 2.50)$

29. En cada caso, determine el valor de la constante c que hace que el enunciado de probabilidad sea correcto.

a. $\Phi(c) = .9838$ **b.** $P(0 \leq Z \leq c) = .291$
c. $P(c \leq Z) = .121$ **d.** $P(-c \leq Z \leq c) = .668$
e. $P(c \leq |Z|) = .016$

30. Encuentre los siguientes percentiles de la distribución normal estándar. Interpole en los casos en que sea apropiado.

a. 91º **b.** 9º **c.** 75º
d. 25º **e.** 6º

31. Determine z_α para lo siguiente

a. $\alpha = .0055$ **b.** $\alpha = .09$
c. $\alpha = .663$

32. Suponga que la fuerza que actúa en una columna que ayuda a soportar un edificio es una variable aleatoria X normalmente distribuida con media de 15.0 kips y desviación estándar de 1.25 kips. Calcule las siguientes probabilidades por estandarización y luego use la tabla A.3

a. $P(X \leq 15)$ **b.** $P(X \leq 17.5)$
c. $P(X \geq 10)$ **d.** $P(14 \leq X \leq 18)$
e. $P(|X - 15| \leq 3)$

33. Las mopeds (motos pequeñas con una cilindrada inferior a 50 cm^3) son muy populares en Europa debido a su movilidad, facilidad de uso y bajo costo. El artículo “Procedure to Verify the Maximum Speed of Automatic Transmission Mopeds in Periodic Motor Vehicle Inspections” (*J. of Automobile Engr.*, 2008: 1615–1623) describió un banco de pruebas rodante para determinar la velocidad máxima del vehículo. Se propone una distribución normal con valor medio de 46.8 km/h y desviación estándar de 1.75 km/h. Considere la posibilidad de seleccionar al azar una sola de esas mopeds.

a. ¿Cuál es la probabilidad de que la velocidad máxima sea a lo sumo 50 km/h?
b. ¿Cuál es la probabilidad de que la velocidad máxima sea al menos de 48 km/h?
c. ¿Cuál es la probabilidad de que la velocidad máxima difiera del valor medio por más de 1.5 desviaciones estándar?

34. El artículo “Reliability of Domestic-Waste Biofilm Reactors” (*J. of Envir. Engr.*, 1995: 785–790) sugiere que la concentración de sustrato (mg/cm^3) del afluente que llega a un reactor está normalmente distribuida con $\mu = .30$ y $\sigma = .06$.

a. ¿Cuál es la probabilidad de que la concentración exceda de .25?
b. ¿Cuál es la probabilidad de que la concentración sea cuando mucho de .10?
c. ¿Cómo caracterizaría el 5% más grande de todos los valores de concentración?

35. Suponga que el diámetro a la altura del pecho (pulg) de árboles de un tipo está normalmente distribuido con $\mu = 8.8$ y $\sigma = 2.8$, como se sugiere en el artículo “Simulating a Harvester-Forwarder Softwood Thinning” (*Forest Products J.*, mayo de 1997; 36–41).

a. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar será por lo menos de 10 pulg? ¿Y que exceda de 10 pulg?
b. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar sea de más de 20 pulg?
c. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar sea de entre 5 y 10 pulg?
d. ¿Qué valor c es tal que el intervalo $(8.8 - c, 8.8 + c)$ incluya 98% de todos los valores de diámetro?
e. Si se seleccionan cuatro árboles al azar, ¿cuál es la probabilidad de que por lo menos uno tenga un diámetro de más de 10 pulg?

36. La dispersión de las atomizaciones de pesticidas es una preocupación constante de los fumigadores y productores agrícolas.

La relación inversa entre el tamaño de gota y el potencial de deriva es bien conocida. El artículo "Effects of 2,4-D Formulation and Quinclorac on Spray Droplet Size and Deposition" (*Weed Technology*, 2005: 1030–1036) investigó los efectos de formulaciones de herbicidas en atomizaciones. Una figura en el artículo sugirió que la distribución normal con media de 1050 μm y desviación estándar de 150 μm fue un modelo razonable de tamaño de gotas de agua (el "tratamiento de control") pulverizada a través de una boquilla de 760 ml/min.

a. ¿Cuál es la probabilidad de que el tamaño de una sola gota sea de menos de 1500 μm? ¿Por lo menos de 1000 μm?

b. ¿Cuál es la probabilidad de que el tamaño de una sola gota sea de entre 1000 y 1500 μm?

c. ¿Cómo caracterizaría el 2% más pequeño de todas las gotas?

d. Si se miden los tamaños de cinco gotas independientemente seleccionadas, ¿cuál es la probabilidad de que por lo menos una exceda de 1500 μm?

37. Suponga que la concentración de cloruro en sangre (mmol/L) tiene una distribución normal con media de 104 y desviación estándar de 5 (información en el artículo "Mathematical Model of Chloride Concentration in Human Blood", *J. of Med. Engr. and Tech.*, 2006; 25–30, incluida una gráfica de probabilidad normal como se describe en la sección 4.6, apoya esta suposición).

a. ¿Cuál es la probabilidad de que la concentración de cloruro sea igual a 105? ¿Sea menor que 105? ¿Sea cuando mucho de 105?

b. ¿Cuál es la probabilidad de que la concentración de cloruro difiera de la media por más de 1 desviación estándar? ¿Depende esta probabilidad de los valores de μ y σ?

c. ¿Cómo caracterizaría el .1% más extremo de los valores de concentración de cloruro?

38. Hay dos máquinas disponibles para cortar corchos para usarse en botellas de vino. La primera produce corchos con diámetros que están normalmente distribuidos con media de 3 cm y desviación estándar de .1 cm. La segunda máquina produce corchos con diámetros que tienen una distribución normal con media de 3.04 cm y desviación estándar de .02 cm. Los corchos aceptables tienen diámetros de entre 2.9 y 3.1 cm. ¿Cuál máquina es más probable que produzca un corcho aceptable?

39. a. Si una distribución normal tiene $\mu = 30$ y $\sigma = 5$, ¿cuál es el 91º percentil de la distribución?

b. ¿Cuál es el 6 percentil de la distribución?

c. El ancho de una línea grabada en un "chip" de circuito integrado normalmente está distribuido con media de 3.000 μm y desviación estándar de .140. ¿Qué valor de ancho separa el 10% de las líneas más anchas del 90% restante?

40. El artículo "Monte Carlo Simulation–Tool for Better Understanding of LRFD" (*J. of Structural Engr.*, 1993: 1586–1599) sugiere que la resistencia a ceder (kg/pulg^2) de un acero grado A36 normalmente está distribuida con $\mu = 43$ y $\sigma = 4.5$.

a. ¿Cuál es la probabilidad de que la resistencia a ceder sea cuando mucho de 40? ¿De más de 60?

b. ¿Qué valor de resistencia a ceder separa al 75% más resistente del resto?

41. El dispositivo de apertura automática de un paracaídas de carga militar se diseñó para que lo abriera a 200 m sobre el suelo. Suponga que la altitud de abertura en realidad tiene una distribución normal con valor medio de 200 m y desviación estándar de 30 m. La carga útil se dañará si el paracaídas se abre a una altitud de menos de 100 m. ¿Cuál es la probabilidad de que se dañe la carga útil de cuando menos uno de cinco paracaídas lanzados en forma independiente?

42. La lectura de temperatura tomada con un termopar colocado en un medio a temperatura constante normalmente está distribuida con media μ, la temperatura real del medio, y desviación estándar σ. ¿Qué valor tendría σ para asegurarse de que el 95% de todas las lecturas están dentro de .1º de μ?

43. Se sabe que la distribución de resistencia de resistores de un tipo es normal y la resistencia del 10% de ellos es mayor de 10.256 ohms y la del 5% es de una resistencia menor de 9.671 ohms. ¿Cuáles son el valor medio y la desviación estándar de la distribución de resistencia?

44. Si la longitud roscada de un perno está normalmente distribuida, ¿cuál es la probabilidad de que la longitud roscada de un perno seleccionado al azar esté

a. dentro de 1.5 desviaciones estándar de su valor medio?

b. a más de 2.5 desviaciones estándar de su valor medio?

c. entre 1 y 2 desviaciones estándar de su valor medio?

45. Una máquina que produce cojinetes de bolas inicialmente se ajustó de modo que el diámetro promedio verdadero de los cojinetes que produce sea de .500 pulg. Un cojinete es aceptable si su diámetro está dentro de .004 pulg de su valor objetivo. Suponga, sin embargo, que el ajuste cambia durante el curso de la producción, de modo que los cojinetes tengan diámetros normalmente distribuidos con valor medio de .499 pulg y desviación estándar de .002 pulg. ¿Qué porcentaje de los cojinetes producidos no será aceptable?

46. La dureza Rockwell de un metal se determina hincando una punta endurecida en la superficie del metal y luego midiendo la profundidad de penetración de la punta. Suponga que la dureza Rockwell de una aleación particular está normalmente distribuida con media de 70 y desviación estándar de 3. (La dureza Rockwell se mide en una escala continua.)

a. Si una probeta es aceptable sólo si su dureza oscila entre 67 y 75, ¿cuál es la probabilidad de que una probeta seleccionada al azar tenga una dureza aceptable?

b. Si el rango de dureza aceptable es $(70 - c, 70 + c)$, ¿con qué valor de c tendría 95% de todas las probetas una dureza aceptable?

c. Si el rango de dureza aceptable es como el del inciso (a) y la dureza de cada una de diez probetas seleccionadas al azar se determina de forma independiente, ¿cuál es el valor esperado de probetas aceptables entre las diez?

d. ¿Cuál es la probabilidad de que cuando mucho ocho de diez probetas independientemente seleccionadas tengan una dureza de menos de 73.84? [*Sugerencia*: $Y =$ el número de entre las diez probetas con dureza de menos de 73.84 es una variable binomial; ¿cuál es p?]

47. La distribución de peso de paquetes enviados de cierta manera es normal con valor medio de 12 lb y desviación estándar de 3.5

lb. El servicio de paquetería desea establecer un valor de peso c más allá del cual habrá un cargo extra. ¿Qué valor de c es tal que 99% de todos los paquetes estén por lo menos 1 lb por debajo del peso de cargo extra?

48. Suponga que la tabla A.3 del apéndice contiene $\Phi(z)$ sólo para $z \geq 0$. Explique cómo aun así podría calcular

a. $P(-1.72 \leq Z \leq -.55)$

b. $P(-1.72 \leq Z \leq .55)$

¿Es necesario tabular $\Phi(z)$ para z negativo? ¿Qué propiedad de la curva normal estándar justifica su respuesta?

49. Considere los bebés nacidos en el rango "normal" de 37–43 semanas de gestación. Datos extensos sustentan la suposición de que el peso al nacer de estos bebés nacidos en Estados Unidos está normalmente distribuido con media de 3432 g y desviación estándar de 482 g. [El artículo "Are Babies Normal?" (*The American Statistician* (1999): 298–302) analizó datos de un año particular; para una selección sensible de intervalos de clase, un histograma no parecía del todo normal pero después de una investigación se determinó que esto se debía a que en algunos hospitales medían el peso en gramos, en otros lo medían a la onza más cercana y luego lo convertían en gramos. Una selección modificada de intervalos de clase que permitía esto produjo un histograma que era descrito muy bien por una distribución normal.]

a. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo exceda de 4000 gramos? ¿Esté entre 3000 y 4000 gramos?

b. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo sea de menos de 2000 gramos o de más de 5000 gramos?

c. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo exceda de 7 libras?

d. ¿Cómo caracterizaría el .1% más extremo de todos los pesos al nacer?

e. Si X es una variable aleatoria con una distribución normal y a es una constante numérica ($a \neq 0$), entonces $Y = aX$ también tiene una distribución normal. Use esto para determinar la distribución de pesos al nacer expresados en libras (forma, media y desviación estándar) y luego calcule otra vez la probabilidad del inciso (c). ¿Cómo se compara ésta con su respuesta previa?

50. En respuesta a preocupaciones sobre el contenido nutricional de las comidas rápidas, McDonald's ha anunciado que utilizará un nuevo aceite de cocinar para sus papas a la francesa que reducirá sustancialmente los niveles de ácidos grasos e incrementará la cantidad de grasa poliinsaturada más benéfica. La compañía afirma que 97 de cada 100 personas no son capaces de detectar una diferencia de sabor entre los nuevos y los viejos aceites. Suponiendo que esta cifra es correcta (como proporción de largo plazo), ¿cuál es la probabilidad aproximada de que en una muestra aleatoria de 1000 individuos que han comprado papas a la francesa en McDonald's,

a. ¿Por lo menos 40 puedan notar la diferencia de sabor entre los dos aceites?

b. Cuando mucho 5% pueda notar la diferencia de sabor entre los dos aceites?

51. La desigualdad de Chebyshev (véase el ejercicio 44 del capítulo 3), es válida para distribuciones continuas y discretas. Estipula que para cualquier número k que satisfaga $k \geq 1$, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$ (véase el ejercicio 44 en el capítulo 3 para una interpretación). Obtenga esta probabilidad en el caso de una distribución normal con $k = 1, 2, 3$ y compare con el límite superior.

52. Sea X el número de defectos en un carrete de cinta magnética de 100 m (una variable de valor entero). Suponga que X tiene aproximadamente una distribución normal con $\mu = 25$ y $\sigma = 5$. Use la corrección de continuidad para calcular la probabilidad de que el número de defectos sea

a. entre 20 y 30, inclusive

b. cuando mucho 30. Menos de 30.

53. Si X tiene una distribución binomial con parámetros $n = 25$ y p, calcule cada una de las siguientes probabilidades mediante la aproximación normal (con la corrección de continuidad) en los casos $p = .5, .6,$ y $.8$ y compare con las probabilidades exactas calculadas con la tabla A.1 del apéndice.

a. $P(15 \leq X \leq 20)$

b. $P(X \leq 15)$

c. $P(20 \leq X)$

54. Suponga que 10% de todas las flechas de acero producidas por medio de un proceso no cumplen con las especificaciones pero pueden ser retrabajadas (en lugar de ser desechadas). Considere una muestra aleatoria de 200 flechas y sea X el número entre éstas que no cumplen con las especificaciones y pueden ser retrabajadas. ¿Cuál es la probabilidad aproximada de que X sea

a. cuando mucho 30?

b. menos que 30?

c. entre 15 y 25 (inclusive)?

55. Suponga que sólo 75% de todos los conductores en un estado usan con regularidad el cinturón de seguridad. Se selecciona una muestra aleatoria de 500 conductores. ¿Cuál es la probabilidad de que

a. entre 360 y 400 (inclusive) de los conductores en la muestra usen con regularidad el cinturón de seguridad?

b. menos de 400 de aquellos en la muestra usen con regularidad el cinturón de seguridad?

56. Demuestre que la relación entre un percentil normal general y el percentil z correspondiente es como se estipuló en esta sección.

57. **a.** Demuestre que si X tiene una distribución normal con parámetros μ y σ, entonces $Y = aX + b$ (una función lineal de X) también tiene una distribución normal. ¿Cuáles son los parámetros de la distribución de Y [es decir, $E(Y)$ y $V(Y)$]? [*Sugerencia*: escriba la función de distribución acumulativa de Y, $P(Y \leq y)$, como una integral que implique la función de densidad de probabilidad de X y luego diferencie con respecto a y para obtener la función de densidad de probabilidad de Y.]

b. Si cuando se mide en °C, la temperatura está normalmente distribuida con media de 115 y desviación estándar de 2, ¿qué se puede decir sobre la distribución de temperatura medida en °F?

58. No existe una fórmula exacta para función de distribución acumulativa normal estándar $\Phi(z)$, aunque se han publicado varias buenas aproximaciones en artículos. La siguiente se tomó de "Approximations for Hand Calculators Using Small Integer Coefficients" (*Mathematics of Computation*, 1977: 214–222). Con $0 < z \le 5.5$,

$$P(Z \ge z) = 1 - \Phi(z)$$
$$\approx .5 \exp\left\{-\left[\frac{(83z + 351)z + 562}{703/z + 165}\right]\right\}$$

El error relativo de esta aproximación es de menos de .042%. Úsela para calcular aproximaciones a las siguientes probabilidades y compare siempre que sea posible con las probabilidades obtenidas con la tabla A.3 del apéndice.

a. $P(Z \ge 1)$ **b.** $P(Z < -3)$
c. $P(-4 < Z < 4)$ **d.** $P(Z > 5)$

4.4 Distribuciones exponencial y gamma

La curva de densidad correspondiente a cualquier distribución normal tiene forma de campana y por consiguiente es simétrica. Existen muchas situaciones prácticas en las cuales la variable de interés para un investigador podría tener una distribución asimétrica. Una familia de distribuciones que tiene esta propiedad es la familia gamma. Primero se considera un caso especial, la distribución exponencial, y luego se le generaliza más adelante en esta sección.

Distribución exponencial

La familia de distribuciones exponenciales proporciona modelos de probabilidad que son muy utilizados en disciplinas de ingeniería y ciencia.

DEFINICIÓN

Se dice que X tiene una **distribución exponencial** con parámetro λ ($\lambda > 0$) si la función de densidad de probabilidad de X es

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{de lo contrario} \end{cases} \tag{4.5}$$

Algunas fuentes escriben la función de densidad de probabilidad exponencial en la forma $(1/\beta)e^{-x/\beta}$, de modo que $\beta = 1/\lambda$. El valor esperado de una variable aleatoria X exponencialmente distribuida es

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx$$

Para obtener este valor esperado se requiere integrar por partes. La varianza de X se calcula utilizando el hecho de que $V(X) = E(X^2) - [E(X)]^2$. La determinación de $E(X^2)$ requiere integrar por partes dos veces en sucesión. Los resultados de estas integraciones son los siguientes:

$$\mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

Tanto la media como la desviación estándar de la distribución exponencial son iguales a $1/\lambda$. En la figura 4.26 aparecen varias gráficas de varias funciones de densidad de probabilidad exponenciales.

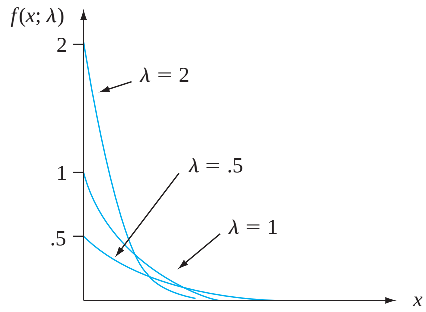


Figura 4.26 Curvas de densidad exponencial

La función de densidad de probabilidad exponencial es fácil de integrar para obtener la función de densidad acumulativa.

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Ejemplo 4.21 El artículo "Probabilistic Fatigue Evaluation of Riveted Railway Bridges" (*J. of Bridge Engr.*, 2008: 237–244) sugirió la distribución exponencial con valor medio de 6 MPa como modelo para la distribución del rango de esfuerzos en las conexiones de determinados puentes. Supongamos que éste es en realidad el verdadero modelo. Entonces $E(X) = 1/\lambda = 6$ implica que $\lambda = .1667$. La probabilidad de que el rango de esfuerzos a lo más sea de 10 MPa es

$$P(X \leq 10) = F(10; .1667) = 1 - e^{-(.1667)(10)} = 1 - .189 = .811$$

La probabilidad de que el rango de esfuerzo sea de entre 5 y 10 MPa es

$$\begin{aligned} P(5 \leq X \leq 10) &= F(10; .1667) - F(5; .1667) = (1 - e^{-1.667}) - (1 - e^{-.8335}) \\ &= .246 \end{aligned}$$

■

La distribución exponencial se utiliza con frecuencia como modelo de la distribución de tiempos entre la ocurrencia de eventos sucesivos, tales como clientes que llegan a una instalación de servicio o llamadas que entran a un conmutador. La razón de esto es que la distribución exponencial está estrechamente relacionada con el proceso de Poisson discutido en el capítulo 3.

PROPOSICIÓN

Suponga que el número de eventos que ocurren en cualquier intervalo de tiempo de duración t tiene una distribución de Poisson con parámetro αt (donde α, la tasa del proceso de eventos, es el número esperado de eventos que ocurren en 1 unidad de tiempo) y que los números de ocurrencias en intervalos no traslapantes son independientes uno de otro. Entonces la distribución del tiempo transcurrido entre la ocurrencia de dos eventos sucesivos es exponencial con parámetro $\lambda = \alpha$.

Aunque una comprobación completa queda fuera del alcance de este libro, el resultado es fácil de verificar para el tiempo X_1 hasta que ocurre el primer evento:

$$\begin{aligned} P(X_1 \leq t) &= 1 - P(X_1 > t) = 1 - P[\text{ningún evento en } (0, t)] \\ &= 1 - \frac{e^{-\alpha t} \cdot (\alpha t)^0}{0!} = 1 - e^{-\alpha t} \end{aligned}$$

la cual es exactamente la función de distribución acumulativa de la distribución exponencial.

Ejemplo 4.22 Suponga que se reciben llamadas durante las 24 horas en una "línea de emergencia para prevención del suicidio" de acuerdo con un proceso de Poisson a razón de $\alpha = .5$ llamadas por día. Entonces el número de días X entre llamadas sucesivas tiene una distribución exponencial con valor de parámetro .5, así que la probabilidad de que transcurran más de dos días entre llamadas es

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2; .5) = e^{-(.5)(2)} = .368$$

El tiempo esperado entre llamadas sucesivas es $1/.5 = 2$ días. ■

Otra aplicación importante de la distribución exponencial es modelar la distribución de la duración de un componente. Una razón parcial de la popularidad de tales aplicaciones es la **propiedad "de no memoria"** de la distribución exponencial. Suponga que la duración de un componente está exponencialmente distribuida con parámetro λ. Después de poner el componente en servicio, se deja que pase un periodo de t_0 horas y luego se ve si el componente sigue trabajando; ¿cuál es ahora la probabilidad de que dure por lo menos t horas más? En símbolos, se desea $P(X \geq t + t_0 | X \geq t_0)$. Por la definición de probabilidad condicional,

$$P(X \geq t + t_0 | X \geq t_0) = \frac{P[(X \geq t + t_0) \cap (X \geq t_0)]}{P(X \geq t_0)}$$

Pero el evento $X \geq t_0$ en el numerador es redundante, puesto que ambos eventos pueden ocurrir si y sólo si $X \geq t + t_0$. Por consiguiente,

$$P(X \geq t + t_0 | X \geq t_0) = \frac{P(X \geq t + t_0)}{P(X \geq t_0)} = \frac{1 - F(t + t_0; \lambda)}{1 - F(t_0; \lambda)} = e^{-\lambda t}$$

Esta probabilidad condicional es idéntica a la probabilidad original $P(X \geq t)$ de que el componente dure t horas. Por lo tanto *la distribución de duración adicional es exactamente la misma que la distribución original de duración*, así que en cada punto en el tiempo el componente no muestra ningún efecto de desgaste. En otras palabras, la distribución de la duración restante es independiente de la antigüedad actual.

Aunque la propiedad de no memoria se justifica por lo menos en forma aproximada en muchos problemas de aplicación, en otras situaciones los componentes se deterioran con el tiempo o de vez en cuando mejoran con él (por lo menos hasta cierto punto). Las distribuciones gamma, de Weibull y lognorma proporcionan modelos de duración más generales (las últimas dos se discuten en la siguiente sección).

La función gamma

Para definir la familia de distribuciones gamma, primero se tiene que introducir una función que desempeña un importante papel en muchas ramas de las matemáticas.

DEFINICIÓN

Con $\alpha > 0$, la **función gamma** $\Gamma(\alpha)$ se define como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx \qquad (4.6)$$

Las propiedades más importantes de la función gamma son las siguientes:

1. Con cualquier $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ [vía integración por partes]

2. Con cualquier entero positivo, n, $\Gamma(n) = (n - 1)!$

3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

De acuerdo con la expresión (4.6), si

$$f(x;\alpha)=\begin{cases}\dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} & x\geq 0\\ 0 & \text{de lo contrario}\end{cases} \tag{4.7}$$

entonces $f(x;\alpha)\geq 0$ y $\int_0^\infty f(x;\alpha)\,dx=\Gamma(\alpha)/\Gamma(\alpha)=1$, así que $f(x;\alpha)$ satisface las dos propiedades básicas de una función de densidad de probabilidad.

La distribución gamma

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución gamma** si la función de densidad de probabilidad de X es

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & x\geq 0\\ 0 & \text{de lo contrario}\end{cases} \tag{4.8}$$

donde los parámetros α y β satisfacen $\alpha>0$, $\beta>0$. La **distribución gamma estándar** tiene $\beta=1$, así que (4.7) da la función de densidad de probabilidad de una variable aleatoria gamma estándar.

La distribución exponencial es resultado de considerar $\alpha=1$ y $\beta=1/\lambda$.

La figura 4.27(a) ilustra las gráficas de la función de densidad de probabilidad gamma $f(x;\alpha,\beta)$ (4.8) para varios pares (α,β), en tanto que la figura 4.27(b) presenta gráficas de la función de densidad de probabilidad gamma estándar. Para la función de densidad de probabilidad estándar cuando $\alpha\leq 1$, $f(x;\alpha)$ es estrictamente decreciente a medida que x se incrementa desde 0; cuando $\alpha>1$, $f(x;\alpha)$ se eleva desde 0 en $x=0$ hasta un máximo y luego decrece. El parámetro β en (4.8) se llama *parámetro de escala* porque los valores diferentes de 1 alargan o comprimen la función de densidad de probabilidad en la dirección x.

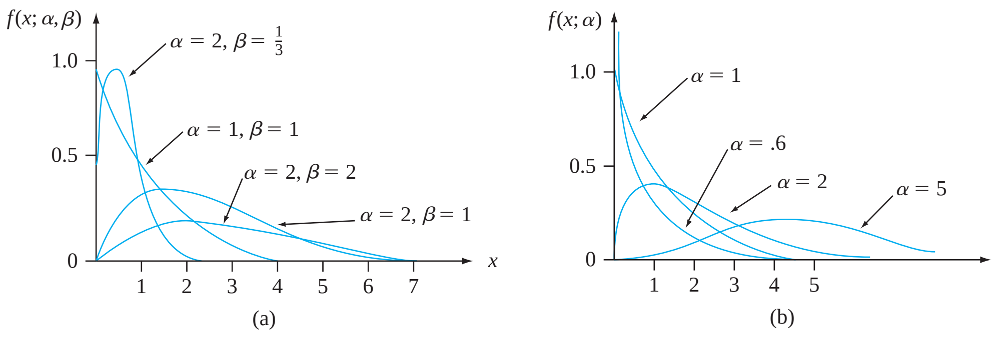


Figura 4.27 (a) Curvas de densidad gamma; (b) Curvas de densidad gamma estándar

La media y la varianza de una variable aleatoria X que tiene la distribución gamma $f(x;\alpha,\beta)$ son

$$E(X)=\mu=\alpha\beta \qquad V(X)=\sigma^2=\alpha\beta^2$$

Cuando X es una variable aleatoria gamma estándar, la función de distribución acumulativa de X,

$$F(x;\alpha)=\int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\,dy \qquad x>0 \tag{4.9}$$

se llama **función gamma incompleta** [en ocasiones la función gamma incompleta se refiere a la expresión (4.9) sin el denominador $\Gamma(\alpha)$ en el integrando]. Existen tablas extensas de $F(x; \alpha)$ disponibles; en la tabla A.4 del apéndice se presenta una pequeña tabulación para $\alpha = 1, 2, \ldots, 10$ y $x = 1, 2, \ldots, 15$.

Ejemplo 4.23 Suponga que el tiempo de reacción X de un individuo seleccionado al azar a un estímulo tiene una distribución gamma estándar con $\alpha = 2$. Como

$$P(a \leq X \leq b) = F(b) - F(a)$$

cuando X es continua,

$$P(3 \leq X \leq 5) = F(5; 2) - F(3; 2) = .960 - .801 = .159$$

La probabilidad de que el tiempo de reacción sea de más de 4 s es

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F(4; 2) = 1 - .908 = .092$$

■

La función gamma incompleta también se utiliza para calcular probabilidades que implican distribuciones gamma no estándar. Estas probabilidades también se obtienen casi instantáneamente con varios paquetes de software.

PROPOSICIÓN

Si X tiene una distribución gamma con parámetros α y β, entonces con cualquier $x > 0$, la función de distribución acumulativa de X es

$$P(X \leq x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

donde $F(\cdot\,; \alpha)$ es la función gamma incompleta.

Ejemplo 4.24 Suponga que el tiempo de sobrevivencia X en semanas de un ratón macho seleccionado al azar expuesto a 240 rads de radiación gamma tiene una distribución gamma con $\alpha = 8$ y $\beta = 15$. (Datos en *Survival Distributions*: *Reliability Applications in the Biomedical Services*, de A. J. Gross y V. Clark, sugiere $\alpha \approx 8.5$ y $\beta \approx 13.3$.) El tiempo de sobrevivencia esperado es $E(X) = (8)(15) = 120$ semanas, en tanto que $V(X) = (8)(15)^2 = 1800$ y $\sigma_X = \sqrt{1800} = 42.43$ semanas. La probabilidad de que un ratón sobreviva entre 60 y 120 semanas es

$$\begin{aligned} P(60 \leq X \leq 120) &= P(X \leq 120) - P(X \leq 60) \\ &= F(120/15; 8) - F(60/15; 8) \\ &= F(8;8) - F(4;8) = .547 - .051 = .496 \end{aligned}$$

La probabilidad de que un ratón sobreviva por lo menos 30 semanas es

$$\begin{aligned} P(X \geq 30) &= 1 - P(X < 30) = 1 - P(X \leq 30) \\ &= 1 - F(30/15; 8) = .999 \end{aligned}$$

■

Distribución ji cuadrada

La distribución ji cuadrada es importante porque es la base de varios procedimientos de inferencia estadística. El papel central desempeñado por la distribución ji cuadrada en inferencia se deriva de su relación con distribuciones normales (véase el ejercicio 71). Se discutirá esta distribución con más detalle en capítulos posteriores.

DEFINICIÓN

Sea ν un entero positivo. Se dice entonces que una variable aleatoria X tiene una **distribución chi cuadrada** con parámetro ν si la función de densidad de probabilidad de X es la densidad gamma con $\alpha = \nu/2$ y $\beta = 2$. La función de densidad de probabilidad de una variable aleatoria ji cuadrada es por lo tanto

$$f(x; \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1}e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{4.10}$$

El parámetro ν se llama **número de grados de libertad** (gl) de X. A menudo se utiliza el símbolo χ^2 en lugar de "ji cuadrada".

EJERCICIOS Sección 4.4 (59–71)

59. Sea X = el tiempo entre dos llegadas sucesivas a la ventanilla de autopago de un banco local. Si X tiene una distribución exponencial con $\lambda = 1$ (la cual es idéntica a una distribución gamma estándar con $\alpha = 1$), calcule lo siguiente:
a. El tiempo esperado entre dos llegadas sucesivas.
b. La desviación estándar del tiempo entre llegadas sucesivas
c. $P(X \leq 4)$ **d.** $P(2 \leq X \leq 5)$

60. Sea X la distancia (m) que un animal recorre desde el sitio de su nacimiento hasta el primer territorio vacante que encuentra. Suponga que para ratas canguro con etiqueta en la cola, X tiene una distribución exponencial con parámetro $\lambda = .01386$ (como lo sugiere el artículo "Competition and Dispersal from Multiple Nests", *Ecology*, 1997: 873–883).
a. ¿Cuál es la probabilidad de que la distancia sea cuando mucho de 100 m? ¿Cuando mucho de 200 m? ¿Entre 100 y 200 m?
b. ¿Cuál es la probabilidad de que la distancia exceda la distancia media por más de 2 desviaciones estándar?
c. ¿Cuál es el valor de la distancia mediana?

61. Los datos recogidos en el Aeropuerto Internacional Toronto Pearson sugiere que una distribución exponencial con valor medio de 2.725 horas es un buen modelo para la duración de la lluvia (*Urban Stormwater Management Planning with Analytical Probabilistic Models*, 2000, p. 69).
a. ¿Cuál es la probabilidad de que la duración de un evento de lluvia en este lugar particular, sea por lo menos 2 horas? ¿A lo más 3 horas? ¿Entre 2 y 3 horas?
b. ¿Cuál es la probabilidad de que la duración de la lluvia supere el valor medio por más de dos desviaciones estándar? ¿Cuál es la probabilidad de que sea menor que el valor medio en más de una desviación estándar?

62. El artículo "Microwave Observations of Daily Antarctic Sea-Ice Edge Expansion and Contribution Rates" *(IEEE Geosci. and Remote Sensing Letters*, 2006: 54–58) establece que "la distribución del avance-retroceso diarios del hielo marino con respecto a cada sensor es similar y es aproximadamente una exponencial doble". La distribución exponencial doble propuesta tiene una función de densidad $f(x) = .5\lambda e^{-\lambda|x|}$ con $-\infty < x < \infty$. La desviación estándar se da como 40.9 km.
a. ¿Cuál es el valor del parámetro λ?
b. ¿Cuál es la probabilidad de que la extensión del cambio del hielo marino esté dentro de 1 desviación estándar del valor medio?

63. Un consumidor está tratando de decidir entre dos planes de llamadas de larga distancia. El primero aplica una sola tarifa de 10¢ por minuto, en tanto que el segundo cobra una tarifa de 99¢ por llamadas hasta de 20 minutos y luego 10¢ por cada minuto adicional que exceda de 20 (suponga que las llamadas que duran un número no entero de minutos son cobradas proporcionalmente a un cargo por minuto entero). Suponga que la distribución de duración de llamadas del consumidor es exponencial con parámetro λ.
a. Explique intuitivamente cómo la selección del plan de llamadas deberá depender de cuál sea la duración de las llamadas.
b. ¿Cuál plan es mejor si la duración esperada de las llamadas es de 10 minutos? ¿Y de 15 minutos? [*Sugerencia*: sea $h_1(x)$ el costo del primer plan cuando la duración de las llamadas es de x minutos y sea $h_2(x)$ la función de costo del segundo plan. Dé expresiones para estas dos funciones de costo y luego determine el costo esperado de cada plan.]

64. Evalúe lo siguiente:
a. $\Gamma(6)$ **b.** $\Gamma(5/2)$
c. $F(4; 5)$ (la función gamma incompleta)
d. $F(5; 4)$ **e.** $F(0; 4)$

65. Si X tiene una distribución gamma estándar con $\alpha = 7$, evalúe lo siguiente:
a. $P(X \leq 5)$ **b.** $P(X < 5)$ **c.** $P(X > 8)$
d. $P(3 \leq X \leq 8)$ **e.** $P(3 < X < 8)$
f. $P(X < 4 \text{ o } X > 6)$

66. Suponga que el tiempo empleado por un estudiante seleccionado al azar que utiliza una terminal conectada a un sistema de computadoras de tiempo compartido tiene una distribución gamma con media de 20 min y varianza de 80 min^2.
a. ¿Cuáles son los valores de α y β?
b. ¿Cuál es la probabilidad de que un estudiante utilice la terminal durante cuando mucho 24 min?
c. ¿Cuál es la probabilidad de que un estudiante utilice la terminal durante entre 20 y 40 min?

67. Suponga que cuando un transistor de cierto tipo se somete a una prueba de duración acelerada, la duración X (en semanas) tiene una distribución gamma con media de 24 semanas y desviación estándar de 12 semanas.

a. ¿Cuál es la probabilidad de que un transistor dure entre 12 y 24 semanas?

b. ¿Cuál es la probabilidad de que un transistor dure cuando mucho 24 semanas? ¿Es la mediana de la distribución de duración menor que 24? ¿Por qué sí o por qué no?

c. ¿Cuál es el 99º percentil de la distribución de duración?

d. Suponga que la prueba termina en realidad después de t semanas. ¿Qué valor de t es tal que sólo el .5% de todos los transistores continuarán funcionando al término?

68. El caso especial de la distribución gamma en la cual α es un entero positivo n se llama distribución de Erlang. Si se reemplaza β por $1/\lambda$ en la expresión (4.8), la función de densidad de probabilidad de Erlang es

$$f(x; \lambda, n) = \begin{cases} \dfrac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Se puede demostrar que si los tiempos entre eventos sucesivos son independientes, cada uno con distribución exponencial con parámetro λ, entonces el tiempo total que transcurre antes de que ocurran los siguientes n eventos tiene una función de densidad de probabilidad $f(x; \lambda, n)$.

a. ¿Cuál es el valor esperado de X? Si el tiempo (en minutos) entre llegadas de clientes sucesivos está exponencialmente distribuido con $\lambda = .5$, ¿cuánto tiempo se puede esperar que transcurra antes de que llegue el décimo cliente?

b. Si el tiempo entre llegadas de clientes está exponencialmente distribuido con $\lambda = .5$, ¿cuál es la probabilidad de que el décimo cliente (después del que acaba de llegar) llegue dentro de los siguientes 30 min?

c. El evento $\{X \leq t\}$ ocurre si al menos ocurren n eventos en las siguientes t unidades de tiempo. Use el hecho de que el número de eventos que ocurren en un intervalo de duración t tiene una distribución de Poisson con parámetro λt para escribir una expresión (que implique probabilidades de Poisson) para la función de distribución acumulativa $F(t; \lambda, n) = P(X \leq t)$.

69. Un sistema consta de cinco componentes idénticos conectados en serie como se muestra:

En cuanto un componente falla, todo el sistema lo hace. Suponga que cada componente tiene una duración que está exponencialmente distribuida con $\lambda = .01$ y que los componentes fallan de manera independiente uno de otro. Defina los eventos $A_i = \{$el componente i-ésimo dura por lo menos t horas$\}$, $i = 1, \ldots, 5$, de modo que los A_i son eventos independientes. Sea $X =$ el tiempo en el cual el sistema falla; es decir, la duración más corta (mínima) entre los cinco componentes.

a. ¿A qué evento equivale el evento $\{X \geq t\}$ que implique $A_1, \ldots, A_5$?

b. Utilizando la independencia de los eventos A_i, calcule $P(X \geq t)$. Luego obtenga $F(t) = P(X \leq t)$ y la función de densidad de probabilidad de X. ¿Qué tipo de distribución tiene X?

c. Suponga que existen n componentes y cada uno tiene una duración exponencial con parámetro λ. ¿Qué tipo de distribución tiene X?

70. Si X tiene una distribución exponencial con parámetro λ, deduzca una expresión general para el $(100p)$º percentil de la distribución. Luego especialícela para obtener la mediana.

71. **a.** ¿A qué evento equivale el evento $\{X^2 \leq y\}$ que implique a la X misma?

b. Si X tiene una distribución normal estándar, use el inciso (a) para escribir la integral que es igual a $P(X^2 \leq y)$. Luego diferénciela con respecto a y para obtener la función de densidad de probabilidad de X^2 [el cuadrado de una variable $N(0, 1)$]. Por último, demuestre que X^2 tiene una distribución ji cuadrada con $\nu = 1$ grado de libertad [véase (4.10)]. [*Sugerencia*: use la siguiente identidad.]

$$\frac{d}{dy}\left\{\int_{a(y)}^{b(y)} f(x)\,dx\right\} = f[b(y)] \cdot b'(y) - f[a(y)] \cdot a'(y)$$

4.5 Otras distribuciones continuas

Las familias de distribuciones normal, gamma (incluida la exponencial) y uniforme proporcionan una amplia variedad de modelos de probabilidad de variables continuas, pero existen muchas situaciones prácticas en las cuales ningún miembro de estas familias se adapta bien a un conjunto de datos observados. Los estadísticos y otros investigadores han desarrollado otras familias de distribuciones que a menudo son apropiadas en la práctica.

Distribución de Weibull

El físico sueco Waloddi Weibull introdujo la familia de distribuciones Weibull en 1939; su artículo de 1951 "A Statistical Distribution Function of Wide Applicability" (*J. of Applied Mechanics*, vol. 18: 293–297) discute varias aplicaciones.

DEFINICIÓN

Se dice que una variable aleatoria X tiene una **distribución de Weibull** con parámetros α y β ($\alpha > 0, \beta > 0$) si la función de densidad de probabilidad de X es

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{4.11}$$

En algunas situaciones, existen justificaciones teóricas para la pertinencia de la distribución de Weibull, pero en muchas aplicaciones $f(x; \alpha, \beta)$ simplemente proporciona una concordancia con los datos observados con valores particulares de α y β. Cuando $\alpha = 1$, la función de densidad de probabilidad se reduce a la distribución exponencial (con $\lambda = 1/\beta$), de modo que la distribución exponencial es un caso especial tanto de la distribución gamma como de la distribución de Weibull. No obstante, existen distribuciones gamma que no son Weibull, y viceversa, por lo que una familia no es un subconjunto de la otra. Tanto α como β pueden ser variadas para obtener diferentes formas de curvas de densidad, como se ilustra en la figura 4.28. β es un parámetro de escala, así que diferentes valores alargan o comprimen la gráfica en la dirección x y α es un parámetro de la forma de la curva.

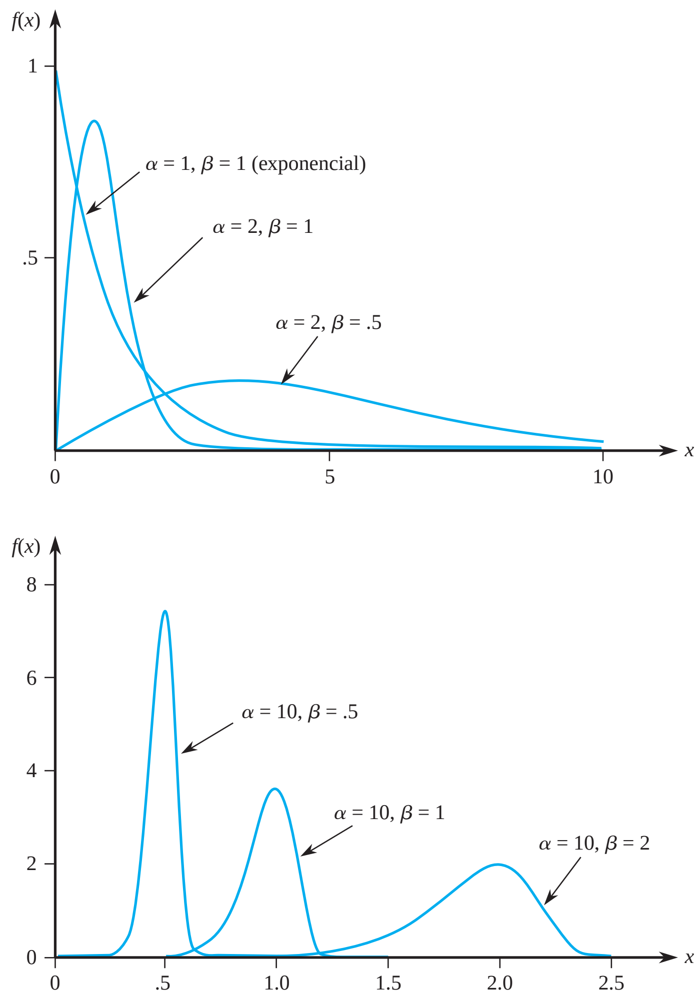


Figura 4.28 Curvas de densidad de Weibull

Si se integra para obtener $E(X)$ y $E(X^2)$ se tiene

$$\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \qquad \sigma^2 = \beta^2\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\}$$

El cálculo de μ y σ^2 requiere por lo tanto el uso de la función gamma.

La integración $\int_0^x f(y;\alpha,\beta)\,dy$ es fácil de realizar para obtener la función de distribución acumulativa de X.

La función de distribución acumulativa de una variable aleatoria de Weibull con parámetros α y β es

$$F(x;\alpha,\beta) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(x/\beta)^\alpha} & x \geq 0 \end{cases} \tag{4.12}$$

Ejemplo 4.25 En años recientes la distribución de Weibull ha sido utilizada para modelar emisiones de varios contaminantes por motores. Sea X la cantidad de emisiones de NO_x (g/gal) de un motor de cuatro tiempos de un tipo seleccionado al azar, y suponga que X tiene una distribución de Weibull con $\alpha = 2$ y $\beta = 10$ (sugeridos por la información que aparece en el artículo "Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors", *J. of the Air and Waste Management Assoc*., 2002: 435–448). La curva de densidad correspondiente se ve exactamente como la de la figura 4.28 con $\alpha = 2$, $\beta = 1$, excepto que ahora los valores 50 y 100 reemplazan a 5 y 10 en el eje horizontal (debido a que β es un "parámetro de escala"). Entonces

$$P(X \leq 10) = F(10; 2, 10) = 1 - e^{-(10/10)^2} = 1 - e^{-1} = .632$$

Asimismo, $P(X \leq 25) = .998$, así que la distribución está concentrada casi por completo en valores entre 0 y 25. El valor c, el cual separa 5% de todos los motores que emiten las más grandes cantidades de NO_x del 95% restante, satisface

$$.95 = 1 - e^{-(c/10)^2}$$

Aislando el término exponencial en un lado, sacando logaritmos y resolviendo la ecuación resultante se obtiene $c \approx 17.3$ como el 95º percentil de la distribución de emisiones. ■

En situaciones prácticas, un modelo de Weibull puede ser razonable excepto que el valor de X más pequeño posible puede ser algún valor γ que no se supuso fuera cero (esto también se aplicaría a un modelo gamma). La cantidad γ puede entonces ser considerada como un tercer (umbral) parámetro de la distribución, lo cual es lo que Weibull hizo en su trabajo original. Con, por ejemplo, $\gamma = 3$, todas las curvas que aparecen en la figura 4.28 se desplazarían 3 unidades a la derecha. Esto equivale a decir que $X - \gamma$ tiene la función de densidad de probabilidad (4.11) de modo que la función de distribución acumulativa de X se obtiene reemplazando x en (4.12) por $x - \gamma$.

Ejemplo 4.26 La comprensión de las propiedades volumétricas del asfalto es importante en el diseño de mezclas que se traducirán en pavimento de alta durabilidad. El artículo "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures?" (*J. of Testing and Evaluation*, sept. 2009: 1–11) utilizó el análisis de algunos datos de la muestra para recomendar que para una mezcla particular, X = volumen vacío de aire (%) se modela con una distribución de Weibull de tres parámetros. Supongamos los valores de los parámetros $\gamma = 4$, $\alpha = 1.3$ y $\beta = .8$ (muy cerca de las estimaciones hechas en el artículo).

Para $x > 4$, la función de distribución acumulativa es

$$F(x;\alpha,\beta,\gamma) = F(x; 1.3, .8, 4) = 1 - e^{-[(x-4)/.8]^{1.3}}$$

La probabilidad de que el volumen vacío de aire de una muestra está entre 5% y 6% es

$$P(5 \le X \le 6) = F(6; 1.3,.8,4) - F(5; 1.3, .8, 4) = e^{-[(5-4)/.8]^{1.3}} - e^{-[(6-4)/.8]^{1.3}}$$
$$= .263 - .037 = .226$$

La figura 4.29 muestra una gráfica de Minitab de la correspondiente función de densidad de Weibull en la que el área sombreada corresponde a la probabilidad que se acaba de calcular.

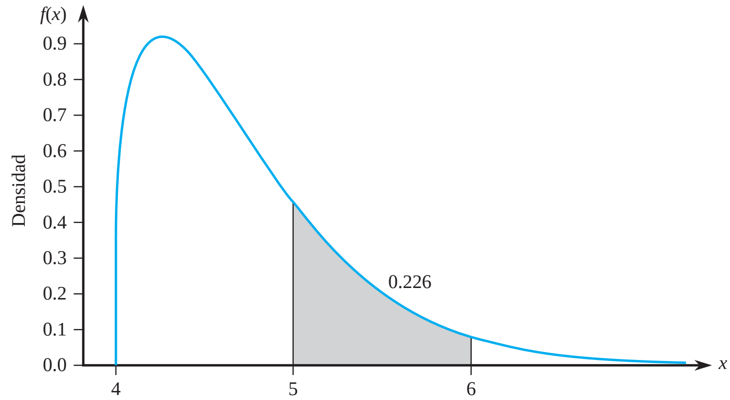


Figura 4.29 Curva de densidad de Weibull con umbral = 4, forma = 1.3, escala = .8 ■

Distribución lognormal

DEFINICIÓN

Se dice que una variable aleatoria no negativa X tiene una **distribución lognormal** si la variable aleatoria $Y = \ln(X)$ tiene una distribución normal. La función de densidad de probabilidad resultante de una variable aleatoria lognormal cuando el $\ln(X)$ está normalmente distribuido con parámetros μ y σ es

$$f(x; \mu, \sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

Hay que tener cuidado aquí; los parámetros μ y σ no son la media y la desviación estándar de X sino de $\ln(X)$. Es común referirse a μ y σ como los parámetros de ubicación y de escala, respectivamente. La media y la varianza de X se puede demostrar que son

$$E(X) = e^{\mu+\sigma^2/2} \qquad V(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$$

En el capítulo 5 se presenta una justificación teórica para esta distribución en conexión con el teorema del límite central, pero como con cualesquiera otras distribuciones, se puede utilizar la lognormal como modelo incluso en la ausencia de semejante justificación. La figura 4.30 ilustra gráficas de la función de densidad de probabilidad lognormal; aunque una curva normal es simétrica, una curva lognormal tiene una asimetría positiva.

Como el $\ln(X)$ tiene una distribución normal, la función de distribución acumulativa de X puede ser expresada en términos de la función de distribución acumulativa $\Phi(z)$ de una variable aleatoria normal estándar Z.

$$F(x; \mu, \sigma) = P(X \le x) = P[\ln(X) \le \ln(x)]$$
$$= P\left(Z \le \frac{\ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \quad x \ge 0 \qquad (4.13)$$

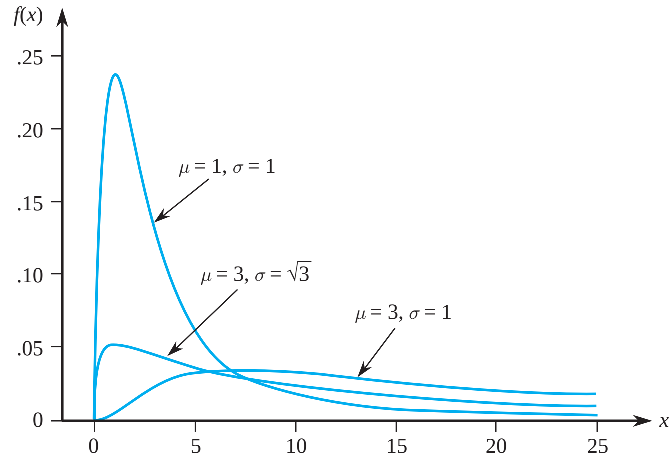


Figura 4.30 Curvas de densidad lognormal

Ejemplo 4.27 De acuerdo con el artículo "Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines" (*Corrosion*, 2009: 332–342), la distribución logarítmica normal ha sido considerada como la mejor opción para describir la distribución de los datos de máxima profundidad de pozo de las tuberías de hierro fundido en el suelo. Los autores sugieren que una distribución logarítmica normal con $\mu = .353$ y $\sigma = .754$ es apropiada para la profundidad de pozo máxima (mm) de tuberías enterradas. Para esta distribución, el valor medio y la varianza de la profundidad del pozo son

$$E(X) = e^{.353+(.754)^2/2} = e^{.6373} = 1.891$$

$$V(X) = e^{2(.353)+(.754)^2} \cdot (e^{(.754)^2} - 1) = (3.57697)(.765645) = 2.7387$$

La probabilidad de que la máxima profundidad de pozo esté entre 1 y 2 mm es

$$P(1 \le X \le 2) = P(\ln(1) \le \ln(X) \le \ln(2)) = P(0 \le \ln(X) \le .693)$$

$$= P\left(\frac{0 - .353}{.754} \le Z \le \frac{.693 - .353}{.754}\right) = \Phi(.47) - \Phi(-.45) = .354$$

Esta probabilidad se ilustra en la figura 4.31 (de Minitab).

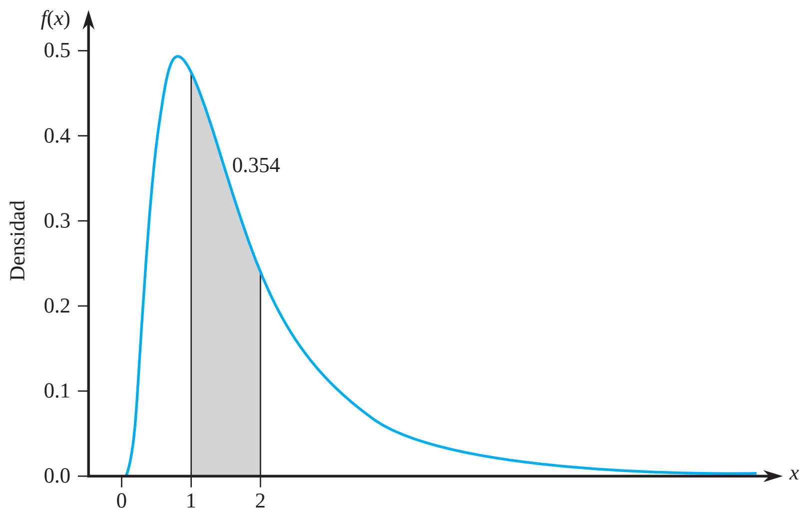


Figura 4.31 Curva de densidad lognormal con ubicación = .353 y escala = .754

¿Qué valor de c es tal que sólo el 1% de todas las muestras tienen una profundidad máxima de pozo que excede c? El valor que se desea es

$$.99 = P(X \leq c) = P\left(Z \leq \frac{\ln(c) - .353}{.754}\right)$$

El valor z crítico 2.33 captura un área de cola superior de .01 ($z_{.01} = 2.33$), y por tanto un área acumulada de .99. Esto implica que

$$\frac{\ln(c) - .353}{.754} = 2.33$$

con lo que $\ln(c) = 2.1098$ y $c = 8.247$. Por lo tanto, 8.247 es el 99º percentil de la distribución de la máxima profundidad de pozo. ■

Distribución beta

Todas las familias de distribuciones continuas estudiadas hasta ahora excepto la distribución uniforme tienen densidad positiva a lo largo de un intervalo infinito (aunque por lo general la función de densidad se reduce con rapidez a cero más allá de unas cuantas desviaciones estándar de la media). La distribución beta proporciona densidad positiva sólo para X en un intervalo de longitud finita.

DEFINICIÓN

Se dice que una variable aleatoria X tiene una **distribución beta** con parámetros α, β (ambos positivos), A y B si la función de densidad de probabilidad de X es

$$f(x; \alpha, \beta, A, B) = \begin{cases} \dfrac{1}{B - A} \cdot \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \left(\dfrac{x - A}{B - A}\right)^{\alpha - 1} \left(\dfrac{B - x}{B - A}\right)^{\beta - 1} & A \leq x \leq B \\ 0 & \text{de lo contrario} \end{cases}$$

El caso $A = 0$, $B = 1$ da la **distribución beta estándar.**

La figura 4.32 ilustra varias funciones de densidad de probabilidad beta estándar. Las gráficas de la función de densidad de probabilidad son similares, excepto que están desplazadas y luego alargadas o comprimidas para ajustarse al intervalo $[A, B]$. A menos que α y β sean enteros, la integración de la función de densidad de probabilidad para calcular probabilidades es difícil. Se deberá utilizar una tabla de la función beta incompleta o un programa de computadora apropiado. La media y varianza de X son

$$\mu = A + (B - A) \cdot \frac{\alpha}{\alpha + \beta} \qquad \sigma^2 = \frac{(B - A)^2 \alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

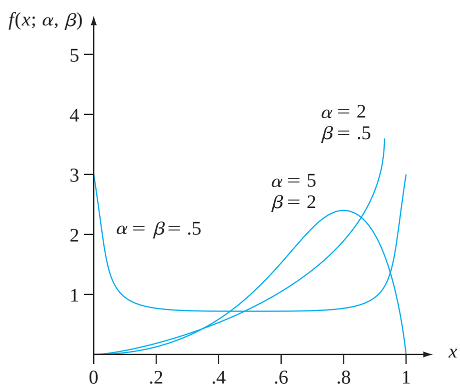


Figura 4.32 Curvas de densidad beta estándar

Ejemplo 4.28 Los gerentes de proyectos a menudo utilizan un método llamado PERT (por las siglas en inglés de técnica de revisión y evaluación de programas) para coordinar las diversas actividades que conforman un gran proyecto. (Una aplicación exitosa ocurrió en la construcción de la nave espacial *Apolo*.) Una suposición estándar en el análisis PERT es que el tiempo necesario para completar cualquier actividad particular una vez que se ha iniciado tiene una distribución beta con A = el tiempo optimista (si todo sale bien) y B = tiempo pesimista (si todo sale mal). Suponga que al construir una casa unifamiliar, el tiempo X (en días) necesario para echar los cimientos tiene una distribución beta con $A = 2$, $B = 5$, $\alpha = 2$ y $\beta = 3$. Entonces, $\alpha/(\alpha + \beta) = .4$, así que $E(X) = 2 + (3)(.4) = 3.2$. Con estos valores de α y β, la función de densidad de probabilidad de X es una función polinomial simple. La probabilidad de que se requieran a lo más tres días para echar los cimientos es

$$P(X \leq 3) = \int_2^3 \frac{1}{3} \cdot \frac{4!}{1!2!}\left(\frac{x-2}{3}\right)\left(\frac{5-x}{3}\right)^2 dx$$

$$= \frac{4}{27}\int_2^3 (x-2)(5-x)^2 dx = \frac{4}{27} \cdot \frac{11}{4} = \frac{11}{27} = .407$$ ■

La distribución beta estándar se utiliza comúnmente para modelar la variación en la proporción o porcentaje de una cantidad que ocurre en diferentes muestras, tal como la proporción de un día de 24 horas que un individuo está despierto o la proporción de cierto elemento en un compuesto químico.

EJERCICIOS Sección 4.5 (72–86)

72. La duración X (en cientos de horas) de un tipo de tubo de vacío tiene una distribución de Weibull con parámetros $\alpha = 2$ y $\beta = 3$. Calcule lo siguiente:

a. $E(X)$ y $V(X)$
b. $P(X \leq 6)$
c. $P(1.5 \leq X \leq 6)$

(Esta distribución de Weibull se sugiere como modelo del tiempo de servicio en "On the Assessment of Equipment Reliability: Trading Data Collection Costs for Precision", *J. of Engr. Manuf.*, 1991: 105–109.)

73. Los autores del artículo "A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses" (*IEEE Trans. on Elect. Insulation*, 1985: 519–522) expresan que "la distribución de Weibull se utiliza mucho en problemas estadísticos relacionados con el desgaste de materiales sólidos aislantes sometidos a envejecimiento y esfuerzo". Proponen el uso de la distribución como modelo del tiempo (en horas) hasta la falla de especímenes aislantes sólidos sometidos a voltaje de CA. Los valores de los parámetros dependen del voltaje y temperatura; suponga $\alpha = 2.5$ y $\beta = 200$ (valores sugeridos por datos que aparecen en el artículo).

a. ¿Cuál es la probabilidad de que la duración de un espécimen sea cuando mucho de 250? ¿De menos de 250? ¿De más de 300?
b. ¿Cuál es la probabilidad de que la duración de un espécimen sea de entre 100 y 250?
c. ¿Qué valor es tal que exactamente 50% de todos los especímenes tengan duraciones que sobrepasen ese valor?

74. Sea X = el tiempo (en 10^{-1} semanas) desde el envío de un producto defectuoso hasta que el cliente lo devuelve. Suponga que el tiempo de devolución mínimo es $\gamma = 3.5$ y que el excedente $X - 3.5$ sobre el mínimo tiene una distribución de Weibull con parámetros $\alpha = 2$ y $\beta = 1.5$ (véase el artículo "Practical Applications of the Weibull Distribution", *Industrial Quality Control*, agosto de 1964: 71–78).

a. ¿Cuál es la función de distribución acumulativa de X?
b. ¿Cuáles son el tiempo de devolución esperado y la varianza del tiempo de devolución? [*Sugerencia*: primero obtenga $E(X - 3.5)$ y $V(X - 3.5)$.]
c. Calcule $P(X > 5)$.
d. Calcule $P(5 \leq X \leq 8)$.

75. Si X tiene una distribución de Weibull con la función de densidad de probabilidad de la expresión (4.11), verifique que $\mu = \beta\Gamma(1 + 1/\alpha)$. [*Sugerencia*: en la integral para $E(X)$ cambie la variable $y = (x/\beta)^\alpha$, de modo que $x = \beta y^{1/\alpha}$.]

76. a. En el ejercicio 72, ¿cuál es la duración mediana de los tubos? [*Sugerencia*: use la expresión (4.12).]
b. En el ejercicio 74, ¿cuál es el tiempo de devolución mediano?
c. Si X tiene una distribución de Weibull con la función de distribución acumulativa de la expresión (4.12), obtenga una expresión general para el percentil $(100p)^{\text{o}}$ de la distribución.
d. En el ejercicio 74, la compañía desea negarse a aceptar devoluciones después de t semanas. ¿Para qué valor de t sólo el 10% de todas las devoluciones serán rechazadas?

77. Los autores del artículo del cual se extrajeron los datos en el ejercicio 1.27 sugirieron que un modelo de probabilidad razonable de la duración de las brocas era una distribución lognormal con $\mu = 4.5$ y $\sigma = .8$.

a. ¿Cuáles son el valor medio y la desviación estándar de la duración?

b. ¿Cuál es la probabilidad de que la duración sea cuando mucho de 100?

c. ¿Cuál es la probabilidad de que la duración sea por lo menos de 200? ¿De más de 200?

78. El artículo "On Assessing the Accuracy of Offshore Wind Turbine Reliability-Based Design Loads from the Environmental Contour Method" (*Intl. J. of Offshore and Polar Engr*., 2005: 132–140) propone la distribución de Weibull con $\alpha = 1.817$ y $\beta = .863$ como modelo de una altura (m) de olas significativa durante 1 hora en un sitio.

a. ¿Cuál es la probabilidad de que la altura de las olas sea cuando mucho de .5 m?

b. ¿Cuál es la probabilidad de que la altura de las olas exceda su valor medio por más de una desviación estándar?

c. ¿Cuál es la mediana de la distribución de la altura de las olas?

d. Para $0 < p < 1$, dé una expresión general para el percentil 100º de la distribución de altura de olas.

79. Cargas de fuentes no puntuales son masas de químicos que viajan al caudal principal de un río y sus afluentes, en flujos que se distribuyen sobre un flujo relativamente de largo alcance, a diferencia de los que entran en puntos bien definidos y regulados. El artículo "Assessing Uncertainty in Mass Balance Calculation of River Nonpoint Source Loads" (*J. of Envir. Engr.*, 2008: 247–258) sugirió que para cierto periodo y lugar, X = carga de fuentes no puntuales de sólidos disueltos totales podría ser modelada con una distribución logarítmica normal con valor medio de 10,281 kg/día/km y un coeficiente de variación $CV = .40$ ($CV = \sigma_X/\mu_X$).

a. ¿Cuáles es el valor medio y la desviación estándar de ln (X)?

b. ¿Cuál es la probabilidad de que X sea a lo sumo 15,000 kg/día/km?

c. ¿Cuál es la probabilidad de que X supere su valor medio y por qué esta probabilidad no es .5?

d. ¿Es 17,000 el percentil 95 de la distribución?

80. **a.** Use la ecuación (4.13) para escribir una fórmula para la mediana $\tilde{\mu}$ de la distribución lognormal. ¿Cuál es la mediana de la distribución de carga del ejercicio 79?

b. Recordando que z_α es la notación para el percentil $100(1 - \alpha)$ de la distribución normal estándar, escriba una expresión para el percentil $100(1 - \alpha)$ de la distribución lognormal. En el ejercicio 79, ¿qué valor excederá la carga recibida sólo 1% del tiempo?

81. Una justificación teórica basada en el mecanismo de falla de cierto material sustenta la suposición de que la resistencia dúctil X de un material tiene una distribución lognormal. Suponga que los parámetros son $\mu = 5$ y $\sigma = .1$.

a. Calcule $E(X)$ y $V(X)$.

b. Calcule $P(X > 125)$.

c. Calcule $P(110 \leq X \leq 125)$.

d. ¿Cuál es el valor de la resistencia dúctil mediana?

e. Si diez muestras diferentes de un acero de aleación de este tipo se sometieran a una prueba de resistencia, ¿cuántas esperaría que tengan una resistencia de por lo menos 125?

f. Si 5% de los valores de resistencia más pequeños fueran inaceptables, ¿cuál sería la resistencia mínima aceptable?

82. El artículo "The Statistics of Phytotoxic Air Pollutants" (*J. of Royal Stat. Soc.*, 1989:183–198) sugiere la distribución lognormal como modelo de la concentración de SO_2 sobre cierto bosque. Suponga que los valores de parámetro son $\mu = 1.9$ y $\sigma = .9$.

a. ¿Cuáles son el valor medio y la desviación estándar de la concentración?

b. ¿Cuál es la probabilidad de que la concentración sea cuando mucho de 10? ¿De entre 5 y 10?

83. ¿Qué condición en relación con α y β es necesaria para que la función de densidad de probabilidad beta estándar sea simétrica?

84. Suponga que la proporción X de área superficial en un cuadrado seleccionado al azar que está cubierto por cierta planta tiene una distribución beta estándar con $\alpha = 5$ y $\beta = 2$.

a. Calcule $E(X)$ y $V(X)$.

b. Calcule $P(X \leq .2)$.

c. Calcule $P(.2 \leq X \leq .4)$.

d. ¿Cuál es la proporción esperada de la región de muestreo no cubierta por la planta?

85. Sea X que tiene una densidad beta estándar con parámetros α y β.

a. Verifique la fórmula para $E(X)$ dada en la sección.

b. Calcule $E[(1 - X)^m]$. Si X representa la proporción de una sustancia compuesta de un ingrediente particular, ¿cuál es la proporción esperada que no se compone de ese ingrediente?

86. Se aplica esfuerzo a una barra de acero de 20 pulg sujeta por cada extremo en una posición fija. Sea Y = la distancia del extremo izquierdo al punto donde se rompe la barra. Suponga que $Y/20$ tiene una distribución beta estándar con $E(Y) = 10$ y $V(Y) = \frac{100}{7}$.

a. ¿Cuáles son los parámetros de la distribución beta estándar pertinente?

b. Calcule $P(8 \leq Y \leq 12)$.

c. Calcule la probabilidad de que la barra se rompa a más de 2 pulg de donde esperaba que se rompiera.

4.6 Gráficas de probabilidad

Un investigador a menudo ha obtenido una muestra numérica $x_1, x_2, \ldots, x_n$ y desea saber si es factible que provenga de una distribución de población de un tipo particular (p. ej., de una distribución normal). Entre otras cosas, muchos procedimientos formales de inferencia estadística están basados en la suposición de que la distribución de población es de un tipo específico. El uso de un procedimiento como ésos es inapropiado si la distribución de probabilidad subyacente real difiere en gran medida del tipo supuesto. Por ejemplo, el artículo "Toothpaste Detergents: A Potential Source of Oral Soft Tissue Damage" (*Intl. J. of Dental Hygiene,* 2008: 193–198) contiene la siguiente declaración: "Debido a que el número de muestras para cada experimento (replicación) fue limitado a tres fuentes según el tipo de tratamiento, se supone que los datos están normalmente distribuidos". Como justificación de este acto de fe, los autores escribieron que "las estadísticas descriptivas mostraron desviaciones estándar que sugieren una distribución normal altamente probable". *Nota*: este argumento no es muy convincente.

Además, el entendimiento de la distribución subyacente en ocasiones puede dar una idea de los mecanismos físicos implicados en la generación de los datos. Una forma efectiva de verificar una suposición distribucional es construir una **gráfica de probabilidad**. La esencia de una gráfica como ésa es que si la distribución en la cual está basada es correcta, los puntos en la gráfica quedarán casi en una línea recta. Si la distribución real es bastante diferente de la utilizada para construir la gráfica, los puntos deberán apartarse sustancialmente de un patrón lineal.

Percentiles muestrales

Los detalles implicados al construir gráficas de probabilidad difieren un poco de una fuente a otra. La base de la construcción es una comparación entre percentiles de los datos muestrales y los percentiles correspondientes de la distribución considerada. Recuérdese que el percentil $(100p)^{o}$ de una distribución continua con función de distribución acumulativa $F(\cdot)$ es el número $\eta(p)$ que satisface $F(\eta(p)) = p$. Es decir, $\eta(p)$ es el número sobre la escala de medición de modo que el área bajo la curva de densidad a la izquierda de $\eta(p)$ es p. Por lo tanto, el percentil 50º, $\eta(.5)$, satisface $F(\eta(.5)) = .5$, y el percentil 90º satisface $F(\eta(.9)) = .9$. Considere como ejemplo la distribución normal estándar, para la cual la función de distribución acumulativa es $\Phi(\cdot)$. En la tabla A.3 del apéndice, el 20º percentil se halla localizando la fila y columna en la cual aparece .2000 (o un número tan cerca de él como sea posible) en el interior de la tabla. Como .2005 aparece en la intersección de la fila $-.8$ y la columna .04, el 20º percentil es aproximadamente $-.84$. Asimismo, el 25º percentil de la distribución normal estándar es (utilizando interpolación lineal) aproximadamente $-.675$.

En general, los percentiles muestrales se definen del mismo modo que los percentiles de una distribución de población. El 50º percentil muestral deberá separar el 50% más pequeño de la muestra del 50% más grande, el 90º percentil deberá ser tal que el 90% de la muestra quede debajo de ese valor y el 10% quede sobre ese valor, y así de manera sucesiva. Desafortunadamente, se presentan problemas cuando en realidad se trata de calcular los percentiles muestrales de una muestra particular de n observaciones. Si, por ejemplo, $n = 10$, se puede separar 20% de estos valores o 30% de los datos, pero no hay ningún valor que separe con exactitud 23% de estas diez observaciones. Para ir más allá, se requiere una definición operacional de percentiles muestrales (éste es un lugar donde diferentes personas hacen cosas un poco diferentes). Recuérdese que cuando n es impar, la mediana muestral o el 50º percentil muestral es el valor medio en la lista ordenada, por ejemplo, el sexto valor más grande cuando $n = 11$. Esto equivale a considerar la observación media como la mitad en la mitad inferior de los datos y la mitad en la mitad superior. Asimismo, supóngase $n = 10$. Entonces, si a este tercer valor más pequeño se le da el nombre de 25º percentil, ese valor se está considerando como la mitad en el grupo inferior (compuesto de las dos observaciones más pequeñas) y la mitad en el grupo superior (las siete observaciones más grandes). Esto conduce a la siguiente definición general de percentiles muestrales.

DEFINICIÓN

Se ordenan las n observaciones muestrales de la más pequeña a la más grande. Entonces la observación i-ésima más pequeña en la lista se considera que es el **$[100(i - .5)/n]^{\circ}$ percentil muestral**.

Una vez que se han calculado los valores porcentuales $100(i - .5)/n$ ($i = 1, 2, \ldots, n$) se pueden obtener los percentiles muestrales correspondientes a porcentajes intermedios mediante interpolación lineal. Por ejemplo, si $n = 10$, los porcentajes correspondientes a las observaciones muestrales ordenadas son $100(1 - .5)/10 = 5\%$, $100(2 - .5)/10 = 15\%$, $25\% \ldots$, y $100(10 - .5)/10 = 95\%$. El 10º percentil está entonces a la mitad entre el 5º percentil (observación muestral más pequeña) y el 15º (segunda observación más pequeña). Para los propósitos de este libro, tal interpolación no es necesaria porque una gráfica de probabilidad se basa sólo en los porcentajes $100(i - .5)/n$ correspondientes a las n observaciones muestrales.

Gráfica de probabilidad

Supóngase ahora que para los porcentajes $100(i - .5)/n$ ($i = 1, \ldots, n$) se determinan los percentiles de una distribución de población especificada cuya factibilidad está siendo investigada. Si la muestra en realidad se seleccionó de la distribución especificada, los percentiles muestrales (observaciones muestrales ordenadas) deberán estar razonablemente próximos a los percentiles de distribución de población correspondientes. Es decir, con $i = 1, 2, \ldots, n$ deberá haber una razonable concordancia entre la i-ésima observación muestral más pequeña y el $[100(i - .5)/n]$º percentil de la distribución especificada. Considérense los pares (percentil poblacional, percentil muestral); es decir, los pares

$$\left(\begin{array}{l} [100(i - .5)/n]^\circ \text{ percentil} \\ \text{de la distribución,} \end{array} \begin{array}{l} i\text{-ésima observación muestral} \\ \text{más pequeña} \end{array} \right)$$

con $i = 1, \ldots, n$. Cada uno de esos pares se grafica como un punto en un sistema de coordenadas bidimensional. Si los percentiles muestrales se acercan a los percentiles de distribución de población correspondientes, el primer número en cada par será aproximadamente igual al segundo número. Los puntos graficados quedarán entonces cerca de una línea a 45°. Desviaciones sustanciales de los puntos graficados con respecto a una línea a 45° hacen dudar de la suposición de que la distribución considerada es la correcta.

Ejemplo 4.29 Un experimentador conoce el valor de cierta constante física. El experimentador realiza $n = 10$ mediciones independientes de este valor por medio de un dispositivo de medición particular y anota los errores de medición resultantes (error = valor observado − valor verdadero). Estas observaciones aparecen en la tabla adjunta.

Porcentaje	5	15	25	35	45
percentil z	−1.645	−1.037	−.675	−.385	−.126
Observación muestral	−1.91	−1.25	−.75	−.53	.20
Porcentaje	55	65	75	85	95
z percentil	.126	.385	.675	1.037	1.645
Observación muestral	.35	.72	.87	1.40	1.56

¿Es factible que el *error de medición* de una variable aleatoria tenga una distribución normal estándar? Los percentiles (z) normales estándar requeridos también se muestran en la tabla. Por lo tanto los puntos en la gráfica de probabilidad son $(-1.645, -1.91)$, $(-1.037, -1.25)$, $\ldots$, y $(1.645, 1.56)$. La figura 4.33 muestra la gráfica resultante. Aunque los puntos se desvían un poco de la línea a 45°, la impresión predominante es que la línea se adapta muy bien a los puntos. La gráfica sugiere que la distribución normal estándar es un modelo de probabilidad razonable para el error de medición.

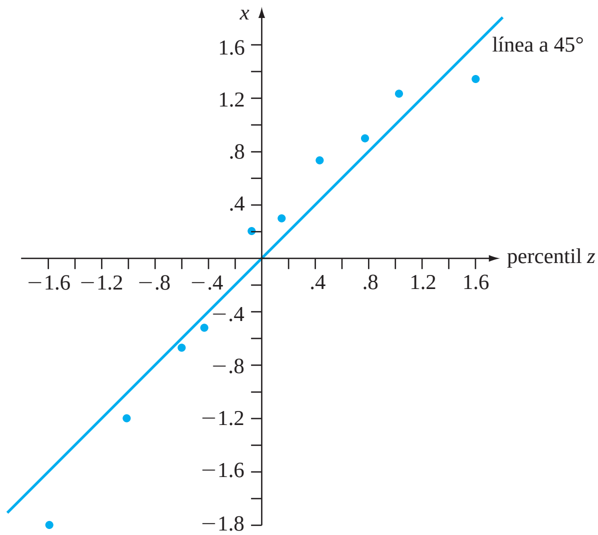


Figura 4.33 Gráficas de pares (percentil z, valor observado) con los datos del ejemplo 4.29; primera muestra

La figura 4.34 muestra una gráfica de pares (percentil z, observación) de una segunda muestra de diez observaciones. La línea a 45° da una buena adaptación a la parte media de la muestra pero no a los extremos. La gráfica tiene apariencia S bien definida. Las dos observaciones muestrales más pequeñas son considerablemente más grandes que los percentiles z correspondientes (los puntos a la extrema izquierda de la gráfica están bien por arriba de la línea a 45°). Asimismo, las dos observaciones muestrales más grandes son mucho más pequeñas que los percentiles z asociados. Esta gráfica indica que la distribución normal estándar no sería una opción factible para el modelo de probabilidad que dio lugar a estos errores de medición observados.

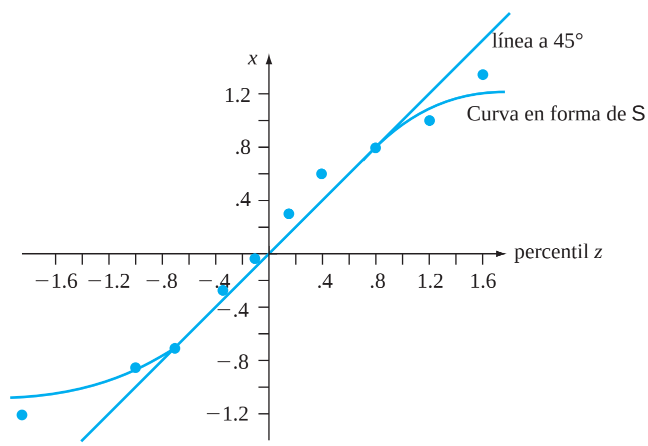


Figura 4.34 Gráficas de pares (percentil z, valor observado) con los datos del ejemplo 4.29; segunda muestra

■

A un investigador en general no le interesa saber con exactitud si una distribución de probabilidad especificada, tal como la distribución normal estándar (normal con $\mu = 0$ y $\sigma = 1$) o la distribución exponencial con $\lambda = .1$, es un modelo plausible de la distribución de población de la cual se seleccionó la muestra. En cambio, la cuestión es si *algún* miembro de una familia de distribuciones de probabilidad especifica un modelo plausible, la familia de distribuciones normales, la familia de distribuciones exponenciales, la familia de distribuciones Weibull, y así sucesivamente. Los valores de los parámetros de una distribución casi nunca se especifican al principio. Si la familia de distribuciones Weibull se considera como modelo de datos de duración, ¿existen *algunos* valores de los parámetros α y β con los cuales la distribución de Weibull correspondiente se adapte bien a los datos? Afortunadamente, casi siempre es el caso de que sólo una gráfica de probabilidad bastará para evaluar la factibilidad de una familia completa. Si la gráfica se desvía sustancialmente de una línea recta, ningún miembro de la familia es factible. Cuando la gráfica es bastante recta, se requiere más trabajo para estimar valores de los parámetros que generen la distribución más razonable del tipo especificado.

Habrá que enfocarse en una gráfica para verificar la normalidad. Tal gráfica es útil en trabajo aplicado porque muchos procedimientos estadísticos formales dan inferencias precisas sólo cuando la distribución de población es por lo menos aproximadamente normal. Estos procedimientos en general no deben ser utilizados si la gráfica de probabilidad normal muestra un alejamiento muy pronunciado de la linealidad. La clave para construir una gráfica de probabilidad normal que comprenda varios elementos es la relación entre los percentiles (z) normales estándar y aquellos de cualquier otra distribución normal:

$$\text{percentil de una distribución normal } (\mu, \sigma) = \mu + \sigma \cdot (\text{percentil } z \text{ correspondiente})$$

Considérese primero el caso, $\mu = 0$. Si cada observación es exactamente igual al percentil normal correspondiente para algún valor de σ, los pares (σ ·[percentil z], observación) quedan sobre una línea a 45°, cuya pendiente es 1. Esto implica que los pares (percentil z, observación) quedan sobre una recta que pasa por (0, 0) (es decir, una con intersección y en 0) pero con pendiente σ en lugar de 1. El efecto del valor no cero de μ es simplemente cambiar la intersección y de 0 a μ.

Una gráfica de los n pares

$$([100(i - .5)/n]^{\text{o}} \text{ percentil } z, \text{ observación } i\text{-ésima más pequeña})$$

en un sistema de coordenadas bidimensional se llama **gráfica de probabilidad normal.** Si las observaciones muestrales se extraen en realidad de una distribución normal con valor medio μ y desviación estándar σ, los puntos deberán quedar cerca de una línea recta con pendiente σ e intersección en μ. Así pues, una gráfica en la cual los puntos quedan cerca de alguna línea recta sugiere que la suposición de una distribución de población normal es factible.

Ejemplo 4.30 La muestra adjunta compuesta de $n = 20$ observaciones de voltaje de ruptura dieléctrica de un pedazo de resina epóxica apareció en el artículo "Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution" (*IEEE Trans. on Dielectrics and Elec. Insul.*, 1996: 43–55). Los valores de $(i - .5)/n$ para los cuales se requieren los percentiles z son $(1 - .5)/20 = .025$, $(2 - .5)/20 = .075, \ldots,$ y .975.

Observación	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
percentil z	−1.96	−1.44	−1.15	−.93	−.76	−.60	−.45	−.32	−.19	−.06
Observación	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
percentil z	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

La figura 4.35 muestra la gráfica de probabilidad normal resultante. La configuración en la gráfica es bastante recta, lo que indica que es factible que la distribución de la población de voltaje de ruptura dieléctrica sea normal.

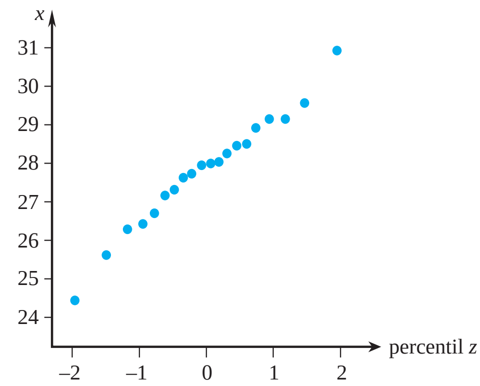


Figura 4.35 Gráfica de probabilidad normal de la muestra de voltaje de ruptura dieléctrica ■

Existe una versión alternativa de una gráfica de probabilidad normal en la cual el eje de los percentiles z es reemplazado por un eje de probabilidad no lineal. La graduación de este eje se construye de modo que los puntos graficados de nuevo queden cerca de una línea cuando la distribución muestreada es normal. La figura 4.36 muestra una gráfica como ésa generada por Minitab con los datos de voltaje de ruptura del ejemplo 4.30.

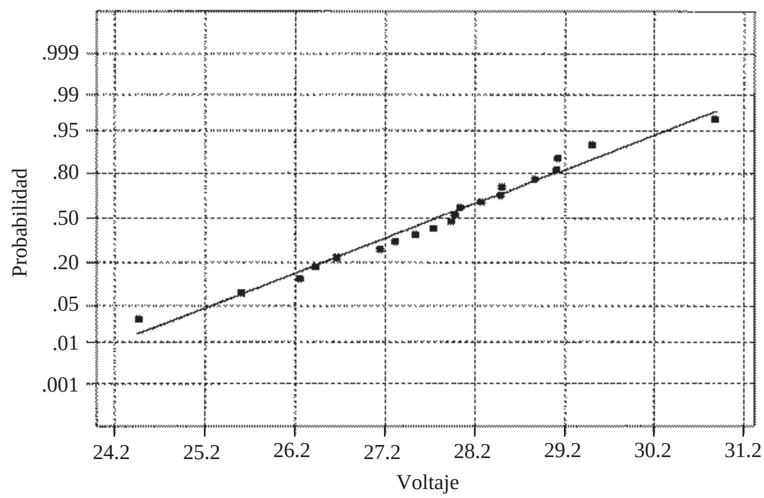


Figura 4.36 Gráfica de probabilidad normal de los datos de voltaje de ruptura generada por Minitab

Una distribución de población no normal a menudo puede ser colocada en una de las siguientes tres categorías:

1. Es simétrica y tiene "colas más livianas" que una distribución normal; es decir, la curva de densidad declina con más rapidez en las colas que una curva normal.
2. Es simétrica y con colas pesadas en comparación con una distribución normal.
3. Es asimétrica.

Una distribución uniforme es de cola liviana, puesto que su función de densidad se reduce a cero afuera de un intervalo finito. La función de densidad $f(x) = 1/[\pi(1 + x^2)]$ para $-\infty < x < \infty$ es de cola pesada, puesto que $1/(1 + x^2)$ declina mucho menos rápidamente que $e^{-x^2/2}$. Las distribuciones lognormal y de Weibull se encuentran entre aquellas que son asimétricas. Cuando los puntos en una gráfica de probabilidad normal no se adhieren a una línea recta, la configuración con frecuencia sugerirá que la distribución de la población se encuentra en una particular de estas tres categorías.

Cuando la distribución de la cual se selecciona la muestra es de cola liviana, las observaciones más grande y más pequeña en general no son tan extremas como podría esperarse de una muestra aleatoria normal. Visualícese una recta trazada a través de la parte media de la gráfica; los puntos a la extrema derecha tienden a estar debajo de la recta (valor observado $<$ percentil z) en tanto que los puntos a la extrema izquierda de la gráfica tienden a quedar sobre la recta (valor observado $>$ percentil z). El resultado es una configuración en forma de S del tipo ilustrado en la figura 4.34.

Una muestra tomada de una distribución de cola pesada también tiende a producir una gráfica en forma de S. Sin embargo, en contraste con el caso de cola liviana, el extremo izquierdo de la gráfica se curva hacia abajo (observado $<$ percentil z), como se muestra en la figura 4.37(a). Si la distribución subyacente es positivamente asimétrica (una cola izquierda corta y una cola derecha larga), las observaciones muestrales más pequeñas serán más grandes que las esperadas de una muestra normal y también lo serán las observaciones más grandes. En este caso, los puntos en ambos extremos de la gráfica quedarán sobre una recta que pasa por la parte media, que produce una configuración curvada, como se ilustra en la figura 4.37(b). Una muestra tomada de una distribución lognormal casi siempre producirá la configuración mencionada. Una gráfica de pares (percentil z, $\ln(x)$) deberá parecerse entonces a una línea recta.

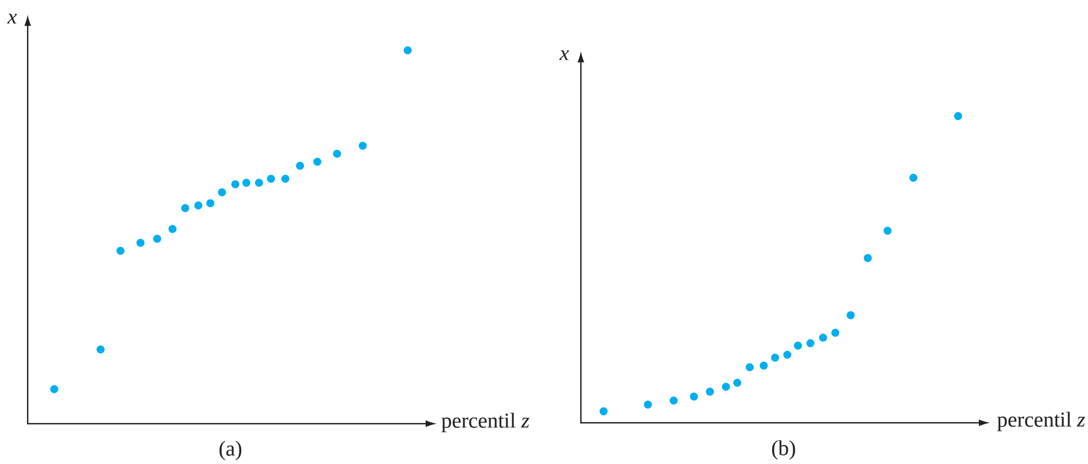


Figura 4.37 Gráficas de probabilidad que sugieren una distribución no normal: (a) una gráfica compatible con una distribución de cola pesada; (b) una gráfica compatible con una distribución positivamente asimétrica

Aunque la distribución de la población sea normal, los percentiles muestrales no coincidirán exactamente con los teóricos debido a la variabilidad del muestreo. ¿Qué tanto pueden desviarse los puntos de la gráfica de probabilidad de un patrón de línea recta antes de que la suposición de normalidad ya no sea plausible? Ésta no es una pregunta fácil de responder. En general, es más probable que una muestra pequeña muestra de una distribución normal produzca una gráfica con un patrón no lineal que una muestra grande. El libro *Fitting Equations to Data* (véase la bibliografía del capítulo 13) presenta los resultados de un estudio de simulación en el cual se seleccionaron numerosas muestras de diferentes tamaños de distribuciones normales. Los autores concluyeron que generalmente varía mucho la apariencia de la gráfica de probabilidad con tamaños de muestra de menos de 30 y sólo con tamaños de muestra mucho más grandes en general predomina el patrón lineal. Cuando una gráfica está basada en un pequeño tamaño de muestra, sólo un alejamiento muy sustancial de la linealidad se deberá considerar como evidencia concluyente de no normalidad. Un comentario similar se aplica a gráficas de probabilidad para comprobar la factibilidad de otros tipos de distribuciones.

Más allá de la normalidad

Considérese una familia de distribuciones de probabilidad que implica dos parámetros θ_1 y θ_2 y sea $F(x; \theta_1, \theta_2)$ la función de distribución acumulativa correspondiente. La familia de distribuciones normales es una esas familias, con $\theta_1 = \mu, \theta_2 = \sigma$ y $F(x; \mu, \sigma) = \Phi[(x - \mu)/\sigma]$. Otro ejemplo es la familia de Weibull, con $\theta_1 = \alpha, \theta_2 = \beta$ y

$$F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$$

Otra familia más de este tipo es la familia gamma, para la cual la función de distribución acumulativa es una integral que implica la función gamma incompleta que no puede ser expresada en alguna forma más simple.

Se dice que los parámetros θ_1 y θ_2 son **parámetros de ubicación** y **escala,** respectivamente, si $F(x; \theta_1, \theta_2)$ es una función de $(x - \theta_1)/\theta_2$. Los parámetros μ y σ de la familia normal son los parámetros de ubicación y escala, respectivamente. Al cambiar μ, la curva de densidad en forma de campana se desplaza a la derecha o izquierda y al cambiar σ se alarga o comprime la escala de medición (la escala sobre el eje horizontal cuando se grafica la función de densidad). La función de distribución acumulativa da otro ejemplo

$$F(x; \theta_1, \theta_2) = 1 - e^{-e^{(x-\theta_1)/\theta_2}} \quad -\infty < x < \infty$$

Se dice que una variable aleatoria con esta función de distribución acumulativa tiene una *distribución de valor extremo*. Se utiliza en aplicaciones que implican la duración de un componente y la resistencia de un material.

Aunque la forma de la función de distribución acumulativa de valor extremo a primera vista pudiera sugerir que θ_1 es el punto de simetría de la función de densidad y por ende la media y la mediana, éste no es el caso. En cambio, $P(X \le \theta_1) = F(\theta_1; \theta_1, \theta_2) = 1 - e^{-1} = .632$ y la función de densidad $f(x; \theta_1, \theta_2) = F'(x; \theta_1, \theta_2)$ es negativamente asimétrica (una larga cola inferior). Asimismo, el parámetro de escala θ_2 no es la desviación estándar ($\mu = \theta_1 - .5772\theta_2$ y $\sigma = 1.283\theta_2$). Sin embargo, al cambiar el valor de θ_1 cambia la ubicación de la curva de densidad, mientras que al cambiar θ_2 cambia la escala del eje de medición.

El parámetro β de la distribución de Weibull es un parámetro de escala, pero α no es un parámetro de ubicación. Un comentario similar es pertinente para los parámetros α y β de la distribución gamma. En la forma usual, la función de densidad de cualquier miembro de la distribución gamma o de Weibull es positiva para $x > 0$ y cero de otra manera. Un parámetro de ubicación puede ser introducido como tercer parámetro γ (se hizo esto para la distribución de Weibull) para desplazar la función de densidad de modo que sea positiva si $x > \gamma$ y cero de otra manera.

Cuando la familia considerada tiene sólo parámetros de ubicación y escala, el tema de si cualquier miembro de la familia es una distribución de población plausible puede ser abordado vía una gráfica de probabilidad única de fácil construcción. Primero se obtienen los percentiles de la *distribución estándar*, una con $\theta_1 = 0$ y $\theta_2 = 1$, con los porcentajes $100(i - .5)/n$ ($i = 1, \ldots, n$). Los n pares (percentil estandarizado, observación) dan los puntos en la gráfica. Esto es exactamente lo que se hizo para obtener una gráfica de probabilidad normal ómnibus. Un tanto sorprendentemente, esta metodología puede ser aplicada para dar una gráfica de probabilidad Weibull más general. El resultado clave es que si X tiene una distribución de Weibull con parámetro de forma α y parámetro de escala β, entonces la variable transformada $\ln(X)$ tiene una distribución de valor extremo con parámetro de ubicación $\theta_1 = \ln(\beta)$ y parámetro de escala $1/\alpha$. Así pues una gráfica de los pares (percentil estandarizado de valor extremo, ln(x)) que muestre un fuerte patrón lineal apoya la selección de la distribución de Weibull como modelo de una población.

Ejemplo 4.31 Las observaciones adjuntas son de la duración (en horas) del aislamiento de aparatos eléctricos cuando la aceleración del esfuerzo térmico y eléctrico se mantuvo fija en valores particulares ("On the Estimation of Life of Power Apparatus Insulation Under Combined Electrical and Thermal Stress", *IEEE Trans. on Electrical Insulation*, 1985: 70–78). Una gráfica de probabilidad de Weibull necesita calcular primero los percentiles 5º, 15º, . . . , y 95º de la distribución de valor extremo estándar. El (100p)º percentil $\eta(p)$ satisface

$$p = F(\eta(p)) = 1 - e^{-e^{\eta(p)}}$$

de donde $\eta(p) = \ln[-\ln(1 - p)]$.

Percentil	−2.97	−1.82	−1.25	−.84	−.51
x	282	501	741	851	1072
ln(*x*)	5.64	6.22	6.61	6.75	6.98
Percentil	−.23	.05	.33	.64	1.10
x	1122	1202	1585	1905	2138
ln(*x*)	7.02	7.09	7.37	7.55	7.67

Los pares (−2.97, 5.64), (−1.82, 6.22), . . . , (1.10, 7.67) se grafican como puntos en la figura 4.38. La derechura de la gráfica argumenta firmemente a favor del uso de la distribución de Weibull como modelo de duración del aislamiento, una conclusión también alcanzada por el autor del citado artículo.

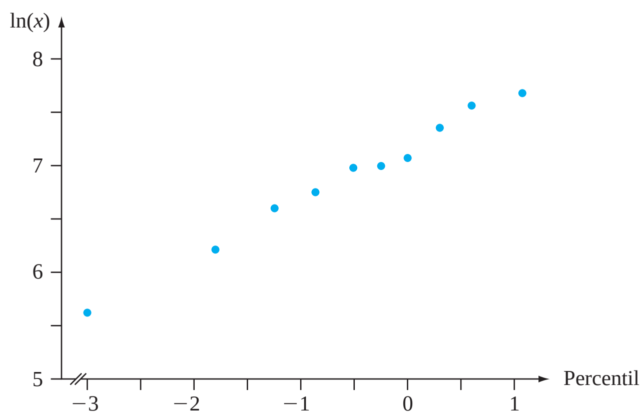
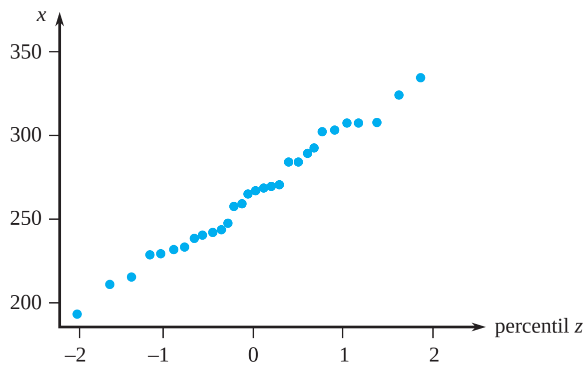


Figura 4.38 Gráfica de probabilidad Weibull de los datos de duración del aislamiento ■

La distribución gamma es un ejemplo de una familia que implica un parámetro de forma para el cual no hay ninguna transformación $h(\cdot)$ de tal suerte que $h(X)$ tenga una distribución que dependa sólo de los parámetros de ubicación y escala. Para construir una gráfica de probabilidad primero se tiene que estimar el parámetro de forma de los datos muestrales (algunos métodos para realizar lo anterior se describen en el capítulo 6). En ocasiones un investigador desea saber si la variable transformada X^θ tiene una distribución normal para algún valor de θ (por convención, $\theta = 0$ es idéntica a la transformación, logarítmica en cuyo caso X tiene una distribución lognormal). El libro *Graphical Methods for Data Analysis*, citado en la bibliografía del capítulo 1, discute este tipo de problema así como también otros refinamientos de construcción de gráficas de probabilidad. Afortunadamente, la amplia disponibilidad de varias gráficas de probabilidad junto con paquetes de software estadísticos significa que el usuario con frecuencia puede evitar los detalles técnicos.

EJERCICIOS Sección 4.6 (87–97)

87. La gráfica de probabilidad normal adjunta se construyó con una muestra de 30 lecturas de tensión de pantallas de malla localizadas detrás de la superficie de tubos de pantallas de video utilizadas en monitores de computadora. ¿Parece factible que la distribución de tensión sea normal?



88. Una muestra de 15 golfistas universitarias fue seleccionada y la velocidad de la cabeza del palo (km/h) mientras se hace *swing* con un *driver* se determinó para cada una, dando como resultado los siguientes datos (“Hip Rotational Velocities During the Full Golf Swing”, *J. of Sports Science and Medicine*, 2009: 296–299):

69.0	69.7	72.7	80.3	81.0
85.0	86.0	86.3	86.7	87.7
89.3	90.7	91.0	92.5	93.0

Los correspondientes percentiles z son

−1.83	−1.28	−0.97	−0.73	−0.52
−0.34	−0.17	0.0	0.17	0.34
0.52	0.73	0.97	1.28	1.83

Construya una gráfica de probabilidad normal y una gráfica de puntos. ¿Es factible que la distribución de la población sea normal?

89. Construya una gráfica de probabilidad normal con la siguiente muestra de observaciones de espesor de recubrimiento de pintura de baja viscosidad (“Achieving a Target Value for a Manufacturing Process: A Case Study”, *J. of Quality Technology*, 1992: 22–26). ¿Se sentiría cómodo estimando el espesor medio de la población con un método que supuso una distribución de población normal?

.83	.88	.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

90. El artículo “A Probabilistic Model of Fracture in Concrete and Size Effects on Fracture Toughness” (*Magazine of Concrete Res.*, 1996: 311–320) da argumentos de por qué la distribución de tenacidad a la fractura en especímenes de concreto debe tener una distribución de Weibull y presenta varios histogramas de datos a los que adaptan bien curvas de Weibull superpuestas. Considere la siguiente muestra de tamaño $n = 18$ observaciones de tenacidad de concreto de alta resistencia (compatible con uno de los histogramas); también se dan los valores de $p_i = (i - .5)/18$.

Observación	.47	.58	.65	.69	.72	.74
p_i	.0278	.0833	.1389	.1944	.2500	.3056
Observación	.77	.79	.80	.81	.82	.84
p_i	.3611	.4167	.4722	.5278	.5833	.6389
Observación	.86	.89	.91	.95	1.01	1.04
p_i	.6944	.7500	.8056	.8611	.9167	.9722

Construya una gráfica de probabilidad de Weibull y coméntela.

91. Construya una gráfica de probabilidad normal con los datos de propagación de grietas por fatiga dados en el ejercicio 39 (capítulo 1). ¿Parece factible que la duración de la propagación tenga una distribución normal? Explique.

92. El artículo “The Load-Life Relationship for M50 Bearings with Silicon Nitride Ceramic Balls” (*Lubrication Engr.*, 1984: 153–159) reporta los datos adjuntos de duración de cojinetes (millones de revoluciones) probados con una carga de 6.45 kN.

47.1	68.1	68.1	90.8	103.6	106.0	115.0
126.0	146.6	229.0	240.0	240.0	278.0	278.0
289.0	289.0	367.0	385.9	392.0	505.0	

a. Construya una gráfica de probabilidad normal. ¿La normalidad es plausible?

b. Construya una gráfica de probabilidad de Weibull. ¿Es adecuada la familia de distribución de Weibull?

93. Construya una gráfica de probabilidad que le permita evaluar qué tan adecuada es la distribución lognormal como modelo de los datos de cantidad de lluvia del ejercicio 83 (capítulo 1).

94. Las observaciones adjuntas son valores de precipitación durante marzo a lo largo de un periodo de 30 años en Minneapolis-St. Paul.

.77	1.20	3.00	1.62	2.81	2.48
1.74	.47	3.09	1.31	1.87	.96
.81	1.43	1.51	.32	1.18	1.89
1.20	3.37	2.10	.59	1.35	.90
1.95	2.20	.52	.81	4.75	2.05

a. Construya e interprete una gráfica de probabilidad normal con este conjunto de datos.
b. Calcule la raíz cuadrada de cada valor y luego construya una gráfica de probabilidad normal basada en estos datos transformados. ¿Parece factible que la raíz cuadrada de la precipitación esté normalmente distribuida?
c. Repita el inciso (b) después de transformar por medio de raíces cúbicas.

95. Use un paquete de software estadístico para construir una gráfica de probabilidad normal de los datos de resistencia última a la tensión dados en el ejercicio 13 del capítulo 1 y comente.

96. Sean $y_1, y_2, \ldots, y_n$, las observaciones muestrales *ordenadas* (con y_1 como la más pequeña y y_n como la más grande). Una verificación sugerida de normalidad es graficar los pares $(\Phi^{-1}((i - .5)/n), y_i)$. Suponga que se cree que las observaciones provienen de una distribución con media 0 y sean $w_1, \ldots, w_n$ los valores *absolutos ordenados* de las x_i. Una gráfica **seminormal** es una gráfica de probabilidad de las w_i. Más específicamente, como $P(|Z| \leq w) = P(-w \leq Z \leq w) = 2\Phi(w) - 1$, una gráfica seminormal es una gráfica de los pares $(\Phi^{-1}\{[(i - .5)/n + 1]/2\}, w_i)$. La virtud de ésta es que los valores apartados pequeños o grandes en la muestra original ahora aparecerán sólo en el extremo superior de la gráfica y no en ambos extremos. Construya una gráfica seminormal con la siguiente muestra de errores de medición, y comente: $-3.78, -1.27, 1.44, -.39, 12.38, -43.40, 1.15, -3.96, -2.34, 30.84$.

97. Las siguientes observaciones de tiempo de falla (miles de horas) se obtuvieron con una prueba de duración acelerada de 16 chips de circuitos integrados de un tipo:

82.8	11.6	359.5	502.5	307.8	179.7
242.0	26.5	244.8	304.3	379.1	212.6
229.9	558.9	366.7	204.6		

Use los percentiles correspondientes de la distribución exponencial con $\lambda = 1$ para construir una gráfica de probabilidad. Luego explique por qué la gráfica valora la aptitud de la muestra habiendo sido generada a partir de *cualquier* distribución exponencial.

EJERCICIOS SUPLEMENTARIOS (98–128)

98. Sea X = el tiempo que una cabeza de lectura-escritura requiere para localizar un registro deseado en un dispositivo de memoria de disco de computadora una vez que la cabeza se ha colocado sobre la pista correcta. Si los discos giran una vez cada 25 milisegundos, una suposición razonable es que X está uniformemente distribuida en el intervalo [0, 25].
a. Calcule $P(10 \leq X \leq 20)$.
b. Calcule $P(X \geq 10)$.
c. Obtenga la función de distribución acumulativa $F(X)$.
d. Calcule $E(X)$ y σ_X.

99. Una barra de 12 pulg que está sujeta por ambos extremos se somete a una cantidad creciente de esfuerzo hasta que se rompe. Sea Y = la distancia del extremo izquierdo al punto donde ocurre la ruptura. Suponga que Y tiene la función de densidad de probabilidad

$$f(y) = \begin{cases} \left(\dfrac{1}{24}\right)y\left(1 - \dfrac{y}{12}\right) & 0 \leq y \leq 12 \\ 0 & \text{de lo contrario} \end{cases}$$

Calcule lo siguiente:
a. La función de distribución acumulativa de Y y grafíquela.
b. $P(Y \leq 4)$, $P(Y > 6)$ y $P(4 \leq Y \leq 6)$.
c. $E(Y)$, $E(Y^2)$ y $V(Y)$.
d. La probabilidad de que el punto de ruptura ocurra a más de 2 pulg del punto de ruptura esperado.
e. La longitud esperada del segmento más corto cuando ocurre la ruptura.

100. Sea X el tiempo hasta la falla (en años) de cierto componente hidráulico. Suponga que la función de densidad de probabilidad de X es $f(x) = 32/(x + 4)^3$ con $x > 0$.
a. Verifique que $f(x)$ es una función de densidad de probabilidad legítima.
b. Determine la función de distribución acumulativa.
c. Use el resultado del inciso (b) para calcular la probabilidad de que el tiempo hasta la falla sea de entre 2 y 5 años.
d. ¿Cuál es el tiempo esperado hasta la falla?
e. Si el componente tiene un valor de recuperación igual a $100/(4 + x)$ cuando su tiempo hasta la falla es x, ¿cuál es el valor de recuperación esperado?

101. El tiempo X para la terminación de cierta tarea tiene una función de distribución acumulativa $F(x)$ dada por

$$\begin{cases} 0 & x < 0 \\ \dfrac{x^3}{3} & 0 \leq x < 1 \\ 1 - \dfrac{1}{2}\left(\dfrac{7}{3} - x\right)\left(\dfrac{7}{4} - \dfrac{3}{4}x\right) & 1 \leq x \leq \dfrac{7}{3} \\ 1 & x > \dfrac{7}{3} \end{cases}$$

a. Obtenga la función de densidad de probabilidad $f(x)$ y trace su gráfica.
b. Calcule $P(.5 \leq X \leq 2)$.
c. Calcule $E(X)$.

102. Se sabe que el voltaje de ruptura de un diodo seleccionado al azar de cierto tipo está normalmente distribuido con valor medio de 40 V y desviación estándar de 1.5 V.

a. ¿Cuál es la probabilidad de que el voltaje de un solo diodo sea de entre 39 y 42?

b. ¿Qué valor es tal que sólo 15% de todos los diodos tienen voltajes que excedan dicho valor?

c. Si se seleccionan cuatro diodos independientemente, ¿cuál es la probabilidad de que por lo menos uno tenga un voltaje de más de 42?

103. El artículo "Computer Assisted Net Weight Control" (*Quality Progress*, 1983: 22–25) sugiere una distribución normal con media de 137.2 oz y desviación estándar de 1.6 oz del contenido real de frascos de cierto tipo. El contenido declarado fue de 135 oz.

a. ¿Cuál es la probabilidad de que un solo frasco contenga más que el contenido declarado?

b. Entre diez frascos seleccionados al azar, ¿cuál es la probabilidad de que por lo menos ocho contengan más que el contenido declarado?

c. Suponiendo que la media permanece en 137.2, ¿a qué valor se tendría que cambiar la desviación estándar de modo que 95% de todos los frascos contengan más que el contenido declarado?

104. Cuando tarjetas de circuito utilizadas en la fabricación de reproductores de discos compactos se someten a prueba, el porcentaje a largo plazo de tarjetas defectuosas es de 5%. Suponga que se recibió un lote de 250 tarjetas y que la condición de cualquier tarjeta particular es independiente de la de cualquier otra.

a. ¿Cuál es la probabilidad aproximada de que por lo menos 10% de las tarjetas en el lote sean defectuosas?

b. ¿Cuál es la probabilidad aproximada de que haya exactamente 10 defectuosas en el lote?

105. El artículo "Characterization of Room Temperature Damping in Aluminum-Indium Alloys" (*Metallurgical Trans.*, 1993: 1611–1619) sugiere que el tamaño de grano de matriz A1 (μm) de una aleación compuesta de 2% de indio podría ser modelado con una distribución normal con valor medio de 96 y desviación estándar de 14.

a. ¿Cuál es la probabilidad de que el tamaño de grano exceda de 100?

b. ¿Cuál es la probabilidad de que el tamaño de grano sea de entre 50 y 80?

c. ¿Qué intervalo (a, b) incluye el 90% central de todos los tamaños de grano (de modo que 5% esté por debajo de a y 5% por encima de b)?

106. El tiempo de reacción (en segundos) a un estímulo es una variable aleatoria continua con función de densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{3}{2} \cdot \dfrac{1}{x^2} & 1 \le x \le 3 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Obtenga la función de distribución acumulativa.

b. ¿Cuál es la probabilidad de que el tiempo de reacción sea cuando mucho de 2.5 s? ¿De entre 1.5 y 2.5 s?

c. Calcule el tiempo de reacción esperado.

d. Calcule la desviación estándar del tiempo de reacción.

e. Si un individuo requiere de más de 1.5 s para reaccionar, una luz se enciende y permanece encendida hasta que transcurre un segundo más o hasta que la persona reacciona (lo que suceda primero). Determine la cantidad de tiempo que se espera que la luz permanezca encendida. [*Sugerencia*: sea $h(X)$ = el tiempo que la luz está encendida como una función del tiempo de reacción X.]

107. Sea X la temperatura a la cual ocurre una reacción química. Suponga que X tiene una función de densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{1}{9}(4 - x^2) & -1 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la gráfica de $f(x)$.

b. Determine la función de distribución acumulativa y grafíquela.

c. ¿Es 0 la temperatura mediana a la cual ocurre la reacción? Si no, ¿es la temperatura mediana más pequeña o más grande que 0?

d. Suponga que esta reacción se realiza independientemente una vez en cada uno de diez laboratorios diferentes y que la función de densidad de probabilidad del tiempo de reacción en cada laboratorio es como se da. Sea Y = el número entre los diez laboratorios en los cuales la temperatura excede de 1. ¿Qué clase de distribución tiene Y? (Dé los nombres y valores de los parámetros.)

108. El artículo "Determination of the MTF of Positive Photoresists Using the Monte Carlo Method" (*Photographic Sci. and Engr.*, 1983: 254–260) propone la distribución exponencial con parámetro $\lambda = .93$ como modelo de la distribución de una longitud de trayectoria libre de fotones (μm) en ciertas circunstancias. Suponga que éste es el modelo correcto.

a. ¿Cuál es la longitud de trayectoria esperada y cuál es la desviación estándar de ésta?

b. ¿Cuál es la probabilidad de que la longitud de trayectoria exceda de 3.0? ¿Cuál es la probabilidad de que la longitud de trayectoria esté entre 1.0 y 3.0?

c. ¿Qué valor es excedido por sólo 10% de todas las longitudes de trayectoria?

109. El artículo "The Prediction of Corrosion by Statistical Analysis of Corrosion Profiles" (*Corrosion Science*, 1985: 305–315) sugiere la siguiente función de distribución acumulativa de la profundidad X del pozo más profundo en un experimento que implica la exposición de acero al carbono manganeso a agua de mar acidificada.

$$F(x; \alpha, \beta) = e^{-e^{-(x-\alpha)/\beta}} \qquad -\infty < x < \infty$$

Los autores proponen los valores $\alpha = 150$ y $\beta = 90$. Suponga que éste es el modelo correcto.

a. ¿Cuál es la probabilidad de que la profundidad del pozo más profundo sea cuando mucho de 150? ¿Cuando mucho 300? ¿De entre 150 y 300?

b. ¿Por debajo de qué valor estará la profundidad del pozo máximo en 90% de todos los experimentos?

c. ¿Cuál es la función de densidad de X?

d. Se puede demostrar que la función de densidad es unimodal (una sola cresta). ¿Por encima de qué valor sobre el eje de medición ocurre esta cresta? (Este valor es la moda.)

e. Se puede demostrar que $E(X) \approx .5772\beta + \alpha$. ¿Cuál es la media de los valores dados de α y β y cómo se compara con la mediana y la moda? Trace la gráfica de la función de densidad. [*Nota*: ésta se conoce como *distribución de valor extremo más grande*.]

110. Sea t = la cantidad de impuesto sobre las ventas que un minorista debe al gobierno por un periodo determinado. El artículo "Statistical Sampling in Tax Audits" (*Statistics and the Law*, 2008: 320–343) se propone modelar la incertidumbre en t, considerándola como una variable aleatoria distribuida normalmente con media μ y la desviación estándar σ (en el artículo, estos dos parámetros se estiman a partir de los resultados de una inspección fiscal que implican n operaciones de muestreo). Si a representa la cantidad a la que el minorista es evaluado, entonces resulta una subevaluación si $t > a$ y una sobreevaluación de resultados, si $a > t$. La función de sanción propuesta (es decir, la pérdida) para la sobre o subevaluación es $L(a, t) = t - a$ si $t > a$ y $= k(a - t)$ si $t \le a$ ($k > 1$ se sugiere para incorporar la idea de que la sobreevaluación es más grave que una subevaluación).

a. Demuestre que $a^* = \mu + \sigma\Phi^{-1}(1/(k + 1))$ es el valor de a que minimiza la pérdida esperada, donde Φ^{-1} es la función inversa de la función de distribución acumulativa normal estándar.

b. Si $k = 2$ (se sugiere en el artículo), $\mu = \$100\,000$, y $\sigma = \$10\,000$, ¿cuál es el valor óptimo de a, y cuál es la probabilidad resultante de la sobreevaluación?

111. La *moda* de una distribución continua es el valor x^* que incrementa al máximo $f(x)$.

a. ¿Cuál es la moda de una distribución normal con parámetros μ y σ?

b. ¿Tiene una sola moda la distribución uniforme con parámetros A y B? ¿Por qué si o por qué no?

c. ¿Cuál es la moda de una distribución exponencial con parámetro λ? (Trace una gráfica.)

d. Si X tiene una distribución gamma con parámetros α y β, y $\alpha > 1$, halle la moda. [*Sugerencia*: $\ln[f(x)]$ se incrementará al máximo si y sólo si $f(x)$ es, y puede ser más simple sacar la derivada de $\ln[f(x)]$.

e. ¿Cuál es la moda de una distribución ji cuadrada con ν grados de libertad?

112. El artículo "Error Distribution in Navigation" (*J. of the Institute of Navigation*, 1971: 429–442) sugiere que una distribución de frecuencia de errores positivos (magnitudes de errores) es mejor aproximada por una distribución exponencial. Sea X = el error de posición lateral (millas náuticas), el cual puede ser positivo o negativo. Suponga que la función de densidad de probabilidad de X es

$$f(x) = (.1)e^{-.2|x|} \quad -\infty < x < \infty$$

a. Trace una gráfica de $f(x)$ y compruebe que $f(x)$ es una función de densidad de probabilidad legítima (demuestre que se integra a 1).

b. Obtenga la función de distribución acumulativa de X y trácela.

c. Calcule $P(X \le 0)$, $P(X \le 2)$, $P(-1 \le X \le 2)$ y la probabilidad de que se cometa un error de más de 2 millas.

113. En algunos sistemas, un cliente es asignado a una o dos prestadoras de servicios. Si el tiempo para que el cliente sea atendido por la prestadora de servicios i tiene una distribución exponencial con parámetro λ_i ($i = 1, 2$) y p es la proporción de todos los clientes atendidos por la prestadora de servicios 1, entonces la función de densidad de probabilidad de X = el tiempo para ser atendido de un cliente seleccionado al azar es

$$f(x; \lambda_1, \lambda_2, p) = \begin{cases} p\lambda_1 e^{-\lambda_1 x} + (1 - p)\lambda_2 e^{-\lambda_2 x} & x \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Ésta a menudo se llama distribución hiperexponencial o distribución exponencial combinada. Esta distribución también se propone como modelo de la cantidad de lluvia en "Modeling Monsoon Affected Rainfall of Pakistan by Point Processes" (*J. of Water Resources Planning and Mgmnt*., 1992: 671-688).

a. Verifique que $f(x; \lambda_1, \lambda_2, p)$ sí es una función de densidad de probabilidad.

b. ¿Cuál es la función de distribución acumulativa $F(x; \lambda_1, \lambda_2, p)$?

c. Si $f(x; \lambda_1, \lambda_2, p)$ es la función de densidad de probabilidad de X, ¿cuál es $E(X)$?

d. Utilizando el hecho de que $E(X^2) = 2/\lambda^2$ cuando X tiene una distribución exponencial con parámetro λ, calcule $E(X^2)$ cuando X tiene la función de densidad de probabilidad $f(x; \lambda_1, \lambda_2, p)$. Luego calcule $V(X)$.

e. El coeficiente de variación de una variable aleatoria (o distribución) es $CV = \sigma/\mu$. ¿Cuál es CV para una variable aleatoria exponencial? ¿Qué puede decir sobre el valor de CV cuando X tiene una distribución hiperexponencial?

f. ¿Cuál es el CV de una distribución Erlang con parámetros λ y n como se definen en el ejercicio 68? [*Nota*: en trabajo aplicado, el CV muestral se utiliza para decidir cuál de las tres distribuciones podría ser apropiada.]

114. Suponga que en un estado particular se permite que los personas físicas que presentan su declaración de impuestos detallen sus deducciones sólo si el total de las deducciones detalladas es por lo menos de \$5000. Sea X (en miles de dólares) el total de deducciones detalladas en un formulario seleccionado al azar. Suponga que X tiene la función de densidad de probabilidad

$$f(x; \alpha) = \begin{cases} k/x^\alpha & x \ge 5 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Encuentre el valor de k. ¿Qué restricción en α es necesaria?

b. ¿Cuál es la función de distribución acumulativa de X?

c. ¿Cuál es la deducción total esperada en un formulario seleccionado al azar? ¿Qué restricción es necesaria en α para que $E(X)$ sea finita?

d. Demuestre que $\ln(X/5)$ tiene una distribución exponencial con parámetro $\alpha - 1$.

115. Sea I_i la corriente de entrada a un transistor e I_0 la corriente de salida. En ese caso la ganancia de corriente es proporcional a $\ln(I_0/I_i)$. Suponga que la constante de proporcionalidad es 1 (lo que conduce a seleccionar una unidad de medición particular), así que la ganancia de corriente $= X = \ln(I_0/I_i)$. Suponga que X está normalmente distribuida con $\mu = 1$ y $\sigma = .05$.

a. ¿Qué tipo de distribución tiene la razón I_0/I_i?

b. ¿Cuál es la probabilidad de que la corriente de salida sea más de dos veces la corriente de entrada?

c. ¿Cuáles son el valor esperado y la varianza de la razón entre corriente de salida y corriente de entrada?

116. El artículo "Response of SiC_f/Si_3N_4 Composites Under Static and Cyclic Loading–An Experimental and Statistical Analysis" (*J. of Engr. Materials and Technology*, 1997: 186–193) sugiere que la resistencia a la tensión (MPa) de compuestos en condiciones especificadas puede ser modelada por una distribución de Weibull con $\alpha = 9$ y $\beta = 180$.

a. Trace una gráfica de la función de densidad.

b. ¿Cuál es la probabilidad de que la resistencia de un espécimen seleccionado al azar exceda de 175? ¿Sea de entre 150 y 175?

c. Si se seleccionan al azar dos especímenes y sus resistencias son independientes entre sí, ¿cuál es la probabilidad de que por lo menos uno tenga una resistencia de entre 150 y 175?

d. ¿Qué valor de resistencia separa al 10% de todos los especímenes más débiles del 90% restante?

117. Si Z tiene una distribución normal estándar, defina una nueva variable aleatoria Y como $Y = \sigma Z + \mu$. Demuestre que Y tiene una distribución normal con parámetros μ y σ. [*Sugerencia*: $Y \le y$ si y sólo si $Z \le$? Use esto para definir la función de distribución acumulativa de Y y luego diferénciela con respecto a y.]

118. a. Suponga que la duración X de un componente, medida en horas, tiene una distribución gamma con parámetros α y β. Sea $Y =$ la duración medida en minutos. Deduzca la función de densidad de probabilidad de Y. [*Sugerencia*: $Y \le y$ si y sólo si $X \le y/60$. Use esto para obtener la función de distribución acumulativa de Y y luego diferénciela para obtener la función de densidad de probabilidad.]

b. Si X tiene una distribución gamma con parámetros α y β, ¿cuál es la distribución de probabilidad de $Y = cX$?

119. En los ejercicios 117 y 118, así como también en muchas otras situaciones, se tiene la función de densidad de probabilidad $f(x)$ de X y se desea conocer la función de densidad de probabilidad de $y = h(x)$. Suponga que $h(\cdot)$ es un función invertible, de modo que $y = h(x)$ se resuelve para x a fin de obtener $x = k(y)$. Entonces se puede demostrar que la función de densidad de probabilidad de Y es

$$g(y) = f[k(y)] \cdot |k'(y)|$$

a. Si X tiene una distribución uniforme con $A = 0$ y $B = 1$, obtenga la función de densidad de probabilidad de $Y = -\ln(X)$.

b. Resuelva el ejercicio 117 utilizando este resultado.

c. Resuelva el ejercicio 118(b) utilizando este resultado.

120. Basado en los datos del experimento de lanzamiento de dardo, el artículo "Shooting Darts" (*Chance*, verano de 1997: 16–19) propuso que los errores horizontales y verticales al apuntar a un blanco deben ser independientes unos de otros, cada uno con una distribución normal con media 0 y varianza σ^2. Se puede demostrar entonces que la distancia V del blanco al punto de aterrizaje es

$$f(v) = \frac{v}{\sigma^2} \cdot e^{-v^2/2\sigma^2} \quad v > 0$$

a. ¿De qué familia introducida en este capítulo es miembro esta función de densidad de probabilidad?

b. Si $\sigma = 20$ mm (cerca del valor sugerido por el artículo), ¿cuál es la probabilidad de que un dardo aterrice dentro de 25 mm (aproximadamente 1 pulg) del blanco?

121. El artículo "Three Sisters Give Birth on the Same Day" (*Chance*, primavera de 2001, 23–25) utilizó el hecho de que tres hermanas de Utah dieron a luz el 11 de marzo de 1998 como base para plantear algunas preguntas interesantes con respecto a coincidencias de fechas de nacimiento.

a. No haciendo caso del año bisiesto y suponiendo que los otros 365 días son igualmente probables, ¿cuál es la probabilidad de que tres nacimientos seleccionados al azar ocurran el 11 de marzo? Asegúrese de indicar qué suposiciones adicionales está haciendo.

b. Con las suposiciones utilizadas en el inciso (a), ¿cuál es la probabilidad de que tres nacimientos seleccionados al azar ocurran el mismo día?

c. El autor sugirió, basado en datos extensos, que el tiempo de gestación (tiempo entre la concepción y el nacimiento) podía ser modelado como si tuviera una distribución normal con valor medio de 280 días y desviación estándar de 19.88 días. Las fechas esperadas para las tres hermanas de Utah fueron el 15 de marzo, el 1 de abril y el 4 de abril, respectivamente. Suponiendo que las tres fechas esperadas están en la media de la distribución, ¿cuál es la probabilidad de que los nacimientos ocurrieran el 11 de marzo? [*Sugerencia*: la desviación de la fecha de nacimiento con respecto a la fecha esperada está normalmente distribuida con media 0.]

d. Explique cómo utilizaría la información del inciso (c) para calcular la probabilidad de una fecha de nacimiento común.

122. Sea X la duración de un componente, con $f(x)$ y $F(x)$ como la función de densidad de probabilidad y la función de distribución acumulativa de X. La probabilidad de que el componente falle en el intervalo $(x, x + \Delta x)$ es aproximadamente $f(x) \cdot \Delta x$. La probabilidad condicional de que falle en $(x, x + \Delta x)$ dado que ha durado por lo menos x es $f(x) \cdot \Delta x/[1 - F(x)]$. Dividiendo esto entre Δx se produce la **función de tasa de falla:**

$$r(x) = \frac{f(x)}{1 - F(x)}$$

Una función de tasa de falla creciente indica que la probabilidad de que los componentes viejos se desgasten es cada vez más grande, mientras que una tasa de falla decreciente evidencia una confiabilidad cada vez más grande con la edad. En la práctica, a menudo se supone una falla "en forma de tina de baño".

a. Si X está exponencialmente distribuida, ¿cuál es $r(x)$?

b. Si X tiene una distribución de Weibull con parámetros α y β, ¿cuál es $r(x)$? ¿Con qué valores de parámetros se incrementará $r(x)$? ¿Con qué valores de parámetros decrecerá $r(x)$ con x?

c. Como $r(x) = -(d/dx)\ln[1 - F(x)]$, $1n[1 - F(x)] = -\int r(x)dx$. Suponga

$$r(x) = \begin{cases} \alpha\left(1 - \dfrac{x}{\beta}\right) & 0 \le x \le \beta \\ 0 & \text{de lo contrario} \end{cases}$$

de modo que si un componente dura β horas, durará por siempre (si bien parece irrazonable, este modelo puede ser utilizado para estudiar el "desgaste inicial"). ¿Cuáles son la función de distribución acumulativa y la función de densidad de probabilidad de X?

123. Sea U que tiene una distribución uniforme en el intervalo [0, 1]. Entonces los valores observados que tienen esta distribución se pueden obtener con un generador de números aleatorios por computadora. Sea $X = -(1/\lambda)\ln(1 - U)$.

a. Demuestre que X tiene una distribución exponencial con parámetro λ. [*Sugerencia*: la función de distribución acumulativa de X es $F(x) = P(X \leq x)$; $X \leq x$ equivale a $U \leq$?]

b. ¿Cómo utilizaría el inciso (a) y un generador de números aleatorios para obtener valores observados derivados de una distribución exponencial con parámetro $\lambda = 10$?

124. Considere una variable aleatoria con media μ y desviación estándar σ, y sea $g(X)$ una función especificada de X. La aproximación de la serie de Taylor de primer orden a $g(X)$ en la cercanía de μ es

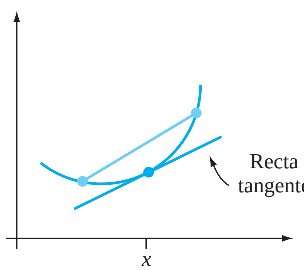
$$g(X) \approx g(\mu) + g'(\mu) \cdot (X - \mu)$$

El miembro del lado derecho de esta ecuación es una función lineal de X. Si la distribución de X está concentrada en un intervalo a lo largo del cual $g(\cdot)$ es aproximadamente lineal [p. ej., $\sqrt{x}$ es aproximadamente lineal en (1, 2)], entonces la ecuación produce aproximaciones a $E(g(X))$ y $V(g(X))$.

a. Dé expresiones para estas aproximaciones. [*Sugerencia*: use reglas de valor esperado y varianza para una función lineal $aX + b$.]

b. Si el voltaje v a través de un medio se mantiene fijo pero la corriente I es aleatoria, entonces la resistencia también será una variable aleatoria relacionada con I por $R = v/I$. Si $\mu_I = 20$ y $\sigma_I = .5$, calcule aproximaciones a μ_R y σ_R.

125. Una función $g(x)$ es *convexa* si la cuerda que conecta dos puntos cualesquiera de su gráfica quedan sobre ésta. Cuando $g(x)$ es diferenciable, una condición equivalente es que para cada x, la línea tangente en x queda por completo sobre o debajo de la gráfica. (Véase la figura a continuación.) ¿Cómo se compara $g(\mu) = g(E(X))$ con $E(g(X))$? [*Sugerencia*: la ecuación de la línea tangente en $x = \mu$ es $y = g(\mu) + g'(\mu) \cdot (x - \mu)$. Use la condición de convexidad, sustituya X por x y considere los valores esperados. [*Nota*: a menos que $g(x)$ sea lineal, la desigualdad resultante (por lo general llamada desigualdad de Jensen) es estricta ($<$ en lugar de $\leq$); es válida tanto con variables aleatorias continuas como discretas.]



126. Si X tiene una distribución de Weibull con parámetros $\alpha = 2$ y β, demuestre que $Y = 2X^2/\beta^2$ tiene una distribución ji cuadrada con $\nu = 2$. [*Sugerencia*: la función de distribución acumulativa de Y es $P(Y \leq y)$; exprese esta probabilidad en la forma $P(X \leq g(y))$, use el hecho de que X tiene una función de distribución acumulativa de la forma de la expresión (4.12) y diferencie con respecto a y para obtener la función de densidad de probabilidad de Y.]

127. El registro crediticio de un individuo es un número calculado basado en el historial crediticio de dicho individuo el cual ayuda a un prestamista a determinar cuánto se le puede prestar o qué límite de crédito debe ser establecido para una tarjeta de crédito. Un artículo en *Los Ángeles Times* presentó datos que sugerían que una distribución beta con parámetros $A = 150$, $B = 850$, $\alpha = 8$, $\beta = 2$ proporcionaría una aproximación razonable a la distribución de registros de crédito estadounidenses. [*Nota*: los registros de crédito son valores enteros].

a. Sea X un registro estadounidense de crédito seleccionado al azar. ¿Cuáles son el valor medio y la desviación estándar de esta variable aleatoria? ¿Cuál es la probabilidad de que X esté dentro de 1 desviación estándar de su valor medio?

b. ¿Cuál es la probabilidad aproximada de que un registro seleccionado al azar excederá de 750 (lo que los prestamistas consideran un muy buen registro)?

128. Sea V el volumen de lluvia y W el volumen de escurrimiento (ambos en mm). De acuerdo con el artículo "Runoff Quality Analysis of Urban Catchments with Analytical Probability Models" (*J. of Water Resource Planning and Management*, 2006: 4–14), el volumen de escurrimiento será 0 si $V \leq \nu_d$ y será $k(V - \nu_d)$ si $V > \nu_d$. Aquí ν_d es el volumen de almacenamiento en una depresión (una constante) y k (también una constante) es el coeficiente de escurrimiento. El artículo citado propone una distribución exponencial con parámetro λ para V.

a. Obtenga una expresión para la función de distribución acumulativa de W. [*Nota*: W no es puramente continua ni puramente discreta; en cambio tiene una distribución "combinada" con un componente discreto en 0 y es continua con valores $w > 0$.]

b. ¿Cuál es la función de densidad de probabilidad de W para $w > 0$? Úsela para obtener una expresión para el valor esperado de volumen de escurrimiento.

Bibliografía

Bury, Karl, *Statistical Distributions in Engineering*, Cambridge Univ. Press, Cambridge, Inglaterra, 1999. Un estudio informativo y fácil de leer de distribuciones y sus propiedades.

Johnson, Norman, Samuel Kotz y N. Balakrishnan, *Continuous Univariate Distributions*, vols. 1–2, Wiley, Nueva York, 1994. Estos dos volúmenes juntos presentan un estudio exhaustivo de varias distribuciones continuas.

Nelson, Wayne, *Applied Data Analysis*, Wiley, Nueva York, 1982. Presenta una amplia discusión de distribuciones y métodos que se utilizan en el análisis de datos de vida útil.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Una buena cobertura de las propiedades generales y distribuciones específicas.

5 Distribuciones de probabilidad conjunta y muestras aleatorias

INTRODUCCIÓN

En los capítulos 3 y 4 se estudiaron modelos de probabilidad para una sola variable aleatoria. Muchos problemas de probabilidad y estadística implican diversas variables aleatorias al mismo tiempo. En este capítulo, primero se discuten modelos de probabilidad del comportamiento conjunto (es decir, simultáneo) de diversas variables aleatorias, con énfasis especial en el caso en el cual las variables son independientes una de otra. En seguida se estudian los valores esperados de funciones de diversas variables aleatorias, incluidas la covarianza y la correlación como medidas del grado de asociación entre dos variables.

Las últimas tres secciones del capítulo consideran funciones de n variables aleatorias $X_1, X_2, \ldots, X_n$, con un enfoque especial en su promedio $(X_1 + \cdots + X_n)/n$. Tal función, por sí misma una variable aleatoria, recibe el nombre de *estadístico.* Se utilizan métodos de probabilidad para obtener información sobre la distribución de un estadístico. El resultado principal de este tipo es el teorema del límite central (TLC), la base de muchos procedimientos inferenciales que implican tamaños de muestra grandes.

5.1 Variables aleatorias conjuntamente distribuidas

Existen muchas situaciones experimentales en las cuales más de una variable aleatoria (rv) será de interés para un investigador. Primero se consideran las distribuciones de probabilidad conjunta para dos variables aleatorias discretas, en seguida para dos variables continuas y por último para más de dos variables.

Dos variables aleatorias discretas

La función de masa de probabilidad (fmp) de una sola variable aleatoria discreta X especifica cuánta masa de probabilidad está colocada en cada valor posible de X. La función de masa de probabilidad conjunta de dos variables aleatorias discretas X y Y describe cuánta masa de probabilidad se coloca en cada posible par de valores (x, y).

DEFINICIÓN

Sean X y Y dos variables aleatorias discretas definidas en el espacio muestral $\mathcal{S}$ de un experimento. La **función de masa de probabilidad conjunta** $p(x, y)$ se define para cada par de números (x, y) como

$$p(x, y) = P(X = x \text{ y } Y = y)$$

Debe ser el caso que $p(x, y) \geq 0$ y $\sum_x \sum_y p(x, y) = 1$.

Ahora sea A cualquier conjunto compuesto de pares de valores (x, y) (p. ej., $A = \{(x, y): x + y = 5\}$ o $\{(x, y): \text{máx}(x, y) \leq 3\}$). Entonces la probabilidad $P[(X, Y) \in A]$ se obtiene sumando la función de masa de probabilidad conjunta incluidos todos los pares en A:

$$P[(X, Y) \in A] = \sum_{(x, y)} \sum_{\in A} p(x, y)$$

Ejemplo 5.1 Una gran agencia de seguros presta servicios a numerosos clientes que han adquirido tanto una póliza de propietario de casa como una póliza de automóvil en la agencia. Por cada tipo de póliza, se debe especificar una cantidad deducible. Para una póliza de automóvil, las opciones son \$100 y \$250, mientras que para la póliza de propietario de casa, las opciones son 0, \$100 y \$200. Suponga que se selecciona al azar un individuo con ambos tipos de póliza de los archivos de la agencia. Sea X = la cantidad deducible sobre la póliza de auto y Y = la cantidad deducible sobre la póliza de propietario de casa. Los posibles pares (X, Y) son entonces (100, 0), (100, 100), (100, 200), (250, 0), (250, 100) y (250, 200); la función de masa de probabilidad conjunta especifica la probabilidad asociada con cada uno de estos pares, con cualquier otro par cuya probabilidad sea cero. Suponga que la **tabla de probabilidad conjunta** siguiente da la función de masa de probabilidad conjunta:

			y	
$p(x, y)$		0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

Entonces $p(100, 100) = P(X = 100 \text{ y } Y = 100) = P(\$100$ deducible sobre ambas pólizas$)$ $= .10$. La probabilidad $P(Y \geq 100)$ se calcula sumando las probabilidades de todos los pares (x, y) para los cuales $y \geq 100$:

$$\begin{aligned} P(Y \geq 100) &= p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) \\ &= .75 \end{aligned}$$ ■

Una vez que la función de masa de probabilidad conjunta de las dos variables X y Y está disponible en principio, es sencillo obtener la distribución de una sola de estas variables. Como ejemplo, sean X y Y el número de los cursos de estadística y de matemáticas, respectivamente, que se están cursando actualmente por una gran estadística seleccionada al azar. Supongamos que se quiere conocer la distribución de X y que cuando $X = 2$, los únicos valores posibles de Y son 0, 1 y 2. Entonces

$$\begin{aligned} p_X(2) = P(X = 2) &= P[(X, Y) = (2, 0) \text{ o } (2, 1) \text{ o } (2, 2)] \\ &= p(2, 0) + p(2, 1) + p(2, 2) \end{aligned}$$

Es decir, la función de masa de probabilidad conjunta se resume en todos los pares de la forma $(2, y)$. De manera más general, para cualquier posible valor x de X, la probabilidad $p_X(x)$ resulta de mantener x fija y sumar la función de masa de probabilidad conjunta $p(x, y)$ sobre toda y para la que el par (x, y) tiene una masa de probabilidad positiva. La misma estrategia se aplica a la obtención de la distribución de Y por sí misma.

DEFINICIÓN

La **función de masa de probabilidad marginal de *X*,** denotada por $p_X(x)$, está dada por

$$p_X(x) = \sum_{y:\, p(x,y)>0} p(x, y) \quad \text{para cada valor posible } x$$

De manera similar, la **función de masa de probabilidad marginal de *Y*** es

$$p_Y(y) = \sum_{x:\, p(x,y)>0} p(x, y) \quad \text{para cada valor posible } y.$$

El uso de la palabra *marginal* aquí es una consecuencia del hecho de que si la función de masa de probabilidad conjunta se muestra en una tabla rectangular como en el ejemplo 5.1, entonces los totales de los renglones dan como resultado la función de masa de probabilidad marginal de X y los totales de las columnas dan como resultado la función de masa de probabilidad marginal de Y. Una vez que estas fmp marginales están disponibles, se puede calcular la probabilidad de cualquier evento que involucre solamente a X o a Y.

Ejemplo 5.2
(Continuación del ejemplo 5.1)

Los valores posibles de X son $x = 100$ y $x = 250$, por lo que si se calculan los totales en los renglones de la tabla de probabilidad conjunta se obtiene

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = .50$$

y

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = .50$$

La función de masa de probabilidad marginal de X es entonces

$$p_X(x) = \begin{cases} .5 & x = 100, 250 \\ 0 & \text{de lo contrario} \end{cases}$$

Asimismo, la función de masa de probabilidad marginal de Y se obtiene con los totales de las columnas como

$$p_Y(y) = \begin{cases} .25 & y = 0, 100 \\ .50 & y = 200 \\ 0 & \text{de lo contrario} \end{cases}$$

Por lo tanto, $P(Y \geq 100) = p_Y(100) + p_Y(200) = .75$ como antes. ■

Dos variables aleatorias continuas

La probabilidad de que el valor observado de una variable aleatoria continua X esté en un conjunto unidimensional A (tal como un intervalo) se obtiene integrando la función de densidad de probabilidad $f(x)$ a lo largo del conjunto A. Asimismo, la probabilidad de que el par (X, Y) de variables aleatorias continuas quede en un conjunto A en dos dimensiones (tal como un rectángulo) se obtiene integrando una función llamada *función de densidad conjunta*.

DEFINICIÓN

Sean X y Y variables aleatorias continuas. Una **función de densidad de probabilidad conjunta** $f(x, y)$ para estas dos variables es una función que satisface $f(x, y) \geq 0$ y $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$. Entonces para cualquier conjunto A en dos dimensiones

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

En particular, si A es el rectángulo bidimensional $\{(x, y): a \leq x \leq b,\ c \leq y \leq d\}$, entonces

$$P[(X, Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

Se puede considerar que $f(x, y)$ especifica una superficie situada a una altura $f(x, y)$ encima del punto (x, y) en un sistema de coordenadas tridimensional. Entonces $P[(X, Y) \in A]$ es el volumen debajo de esta superficie y sobre la región A, similar al área bajo una curva en el caso de una sola variable aleatoria. Esto se ilustra en la figura 5.1.

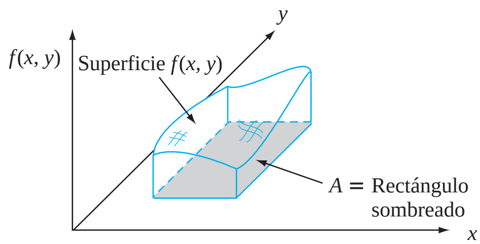


Figura 5.1 $P[(X, Y) \in A]$ = volumen bajo la superficie de densidad sobre A

Ejemplo 5.3 Un banco dispone tanto de una ventanilla para automovilistas como de una ventanilla normal. En un día seleccionado al azar, sea X = la proporción de tiempo que la ventanilla para automovilistas está en uso (por lo menos un cliente está siendo atendido o está esperando ser atendido) y Y = la proporción del tiempo que la ventanilla normal está en uso. Entonces el conjunto de valores posibles de (X, Y) es el rectángulo $D = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Suponga que la función de densidad de probabilidad conjunta de (X, Y) está dada por

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Para verificar que ésta es una función de densidad de probabilidad legítima, obsérvese que $f(x, y) \geq 0$ y

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy &= \int_0^1\int_0^1 \frac{6}{5}(x + y^2)\, dx\, dy \\ &= \int_0^1\int_0^1 \frac{6}{5}x\, dx\, dy + \int_0^1\int_0^1 \frac{6}{5}y^2\, dx\, dy \\ &= \int_0^1 \frac{6}{5}x\, dx + \int_0^1 \frac{6}{5}y^2\, dy = \frac{6}{10} + \frac{6}{15} = 1\end{aligned}$$

La probabilidad de que ninguna ventanilla esté ocupada más de un cuarto del tiempo es

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)\,dx\,dy$$
$$= \frac{6}{5}\int_0^{1/4}\int_0^{1/4} x\,dx\,dy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2\,dx\,dy$$
$$= \frac{6}{20} \cdot \left.\frac{x^2}{2}\right|_{x=0}^{x=1/4} + \frac{6}{20} \cdot \left.\frac{y^3}{3}\right|_{y=0}^{y=1/4} = \frac{7}{640}$$
$$= .0109$$

■

La función de densidad de probabilidad marginal de cada variable se puede obtener de una manera análoga a lo que se hizo en el caso de dos variables discretas. La función de densidad de probabilidad marginal de X en el valor x resulta de mantener x fija en el par (x, y) e *integrando* la función de densidad de probabilidad conjunta sobre y. La integración de la función de densidad de probabilidad conjunta con respecto a x da como resultado la función de densidad de probabilidad marginal de Y.

DEFINICIÓN

Las **funciones de densidad de probabilidad marginal** de X y Y, denotadas por $f_X(x)$ y $f_Y(y)$, respectivamente, están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \text{para } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \quad \text{para } -\infty < y < \infty$$

Ejemplo 5.4 (Continuación del ejemplo 5.3)

La función de densidad de probabilidad marginal de X, la cual da la distribución de probabilidad del tiempo que permanece ocupada la ventanilla para automovilistas sin referencia a la ventanilla normal, es

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 \frac{6}{5}(x + y^2)\,dy = \frac{6}{5}x + \frac{2}{5}$$

con $0 \le x \le 1$ y 0 de lo contrario. La función de densidad de probabilidad marginal de Y es

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \le y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Entonces

$$P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right) = \int_{1/4}^{3/4} f_Y(y)\,dy = \frac{37}{80} = .4625$$

■

En el ejemplo 5.3, la región de densidad conjunta positiva fue un rectángulo, lo que facilitó el cálculo de las funciones de densidad de probabilidad marginal. Considere ahora un ejemplo en el cual la región de densidad positiva es más complicada.

Ejemplo 5.5

Una compañía de nueces comercializa latas de nueces combinadas de lujo que contienen almendras, nueces de acajú y cacahuates. Suponga que el peso neto de cada lata es exactamente 1 lb, pero la contribución al peso de cada tipo de nuez es aleatoria. Como los tres pesos suman 1, un modelo de probabilidad conjunta de dos cualesquiera da toda la información necesaria sobre el peso del tercer tipo. Sea X = el peso de las almendras en una lata seleccionada y Y = el peso de las nueces de acajú. Entonces la región de densidad positiva es $D = \{(x, y): 0 \le x \le 1, 0 \le y \le 1, x + y \le 1\}$, la región sombreada ilustrada en la figura 5.2.

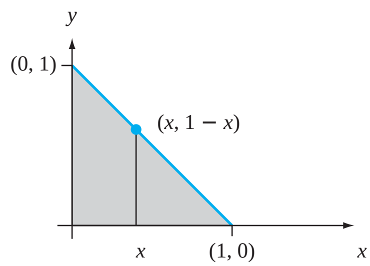


Figura 5.2 Región de densidad positiva para el ejemplo 5.5

Ahora sea la función de densidad de probabilidad conjunta de (X, Y)

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Con cualquier x fija, $f(x, y)$ se incrementa con y; con y fija, $f(x, y)$ se incrementa con x. Esto es apropiado porque las palabras *de lujo* implican que la mayor parte de lata deberá estar compuesta de almendras y nueces de acajú en lugar de cacahuates, así que la función de densidad deberá ser grande cerca del límite superior y pequeña cerca del origen. La superficie determinada por $f(x, y)$ se inclina hacia arriba desde cero a medida que (x, y) se alejan de uno u otro ejes.

Claramente, $f(x, y) \ge 0$. Para verificar la segunda condición sobre una función de densidad de probabilidad conjunta, recuérdese que una integral doble se calcula como una integral iterada manteniendo una variable fija (tal como x en la figura 5.2), integrando con los valores de la otra variable localizados a lo largo de la línea recta que pasa a través del valor de la variable fija y finalmente integrando todos los valores posibles de la variable fija. Así pues

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dy\,dx = \iint_D f(x, y)\,dy\,dx = \int_0^1 \left\{\int_0^{1-x} 24xy\,dy\right\} dx$$
$$= \int_0^1 24x\left\{\frac{y^2}{2}\bigg|_{y=0}^{y=1-x}\right\} dx = \int_0^1 12x(1-x)^2\,dx = 1$$

Para calcular la probabilidad de que los dos tipos de nueces conformen cuando mucho 50% de la lata, sea $A = \{(x, y)\colon 0 \le x \le 1, 0 \le y \le 1, \text{ y } x + y \le .5\}$, como se muestra en la figura 5.3. Entonces

$$P((X, Y) \in A) = \iint_A f(x, y)\,dx\,dy = \int_0^{.5}\int_0^{.5-x} 24xy\,dy\,dx = .0625$$

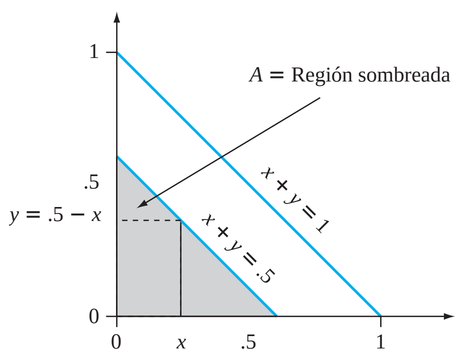


Figura 5.3 Cálculo de $P[(X, Y) \in A]$ para el ejemplo 5.5

La función de densidad de probabilidad marginal de las almendras se obtiene manteniendo X fija en x e integrando la función de densidad de probabilidad conjunta $f(x, y)$ a lo largo de la línea vertical que pasa por x:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \begin{cases} \displaystyle\int_0^{1-x} 24xy\, dy = 12x(1 - x)^2 & 0 \leq x \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Por simetría de $f(x, y)$ y la región D, la función de densidad de probabilidad marginal de Y se obtiene reemplazando x y X en $f_X(x)$ por y y Y, respectivamente. ■

Variables aleatorias independientes

En muchas situaciones, la información sobre el valor observado de una de las dos variables X y Y da información sobre el valor de la otra variable. En el ejemplo 5.1, la probabilidad marginal de X con $x = 250$ fue de .5, como lo fue la probabilidad de que $X = 100$. Si, no obstante, se dice que el individuo seleccionado tuvo $Y = 0$, entonces $X = 100$ es cuatro veces más probable que $X = 250$. Por lo tanto existe dependencia entre las dos variables.

En el capítulo 2 se señaló que una forma de definir la independencia de dos eventos es vía la condición de que $P(A \cap B) = P(A) \cdot P(B)$. A continuación se da una definición análoga de la independencia de dos variables aleatorias.

DEFINICIÓN

Se dice que dos variables aleatorias X y Y son **independientes** si por cada par de valores x y y

$$p(x, y) = p_X(x) \cdot p_Y(y) \qquad \text{cuando } X \text{ y } Y \text{ son discretas}$$

o (5.1)

$$f(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{cuando } X \text{ y } Y \text{ son continuas}$$

Si (5.1) no se satisface con todos los pares (x, y), entonces se dice que X y Y son **dependientes.**

La definición dice que dos variables son independientes si su función de masa de probabilidad conjunta o función de densidad de probabilidad conjunta es el producto de las dos funciones de masa de probabilidad marginales o de las funciones de densidad de probabilidad marginales. Intuitivamente, la independencia dice que conocer el valor de una de las variables no proporciona información adicional acerca de cuál puede ser el valor de la otra variable.

Ejemplo 5.6 En la situación de la agencia de seguros de los ejemplos 5.1 y 5.2,

$$p(100, 100) = .10 \neq (.5)(.25) = p_X(100) \cdot p_Y(100)$$

de modo que X y Y no son independientes. La independencia de X y Y requiere que *toda* entrada en la tabla de probabilidad conjunta sea el producto de las probabilidades marginales que aparecen en los renglones y columnas correspondientes. ■

Ejemplo 5.7 (Continuación del ejemplo 5.5) Como $f(x, y)$ tiene la forma de un producto, X y Y parecerían ser independientes. Sin embargo, aunque $f_X\left(\frac{3}{4}\right) = f_Y\left(\frac{3}{4}\right) = \frac{9}{16}$, $f\left(\frac{3}{4}, \frac{3}{4}\right) = 0 \neq \frac{9}{16} \cdot \frac{9}{16}$ de modo que las variables no son en realidad independientes. Para que sean independientes $f(x, y)$ debe tener la forma $g(x) \cdot h(y)$ y la región de densidad positiva debe ser un rectángulo con sus lados paralelos a los ejes de coordenadas. ■

La independencia de dos variables aleatorias es más útil cuando la descripción del experimento en estudio sugiere que X y Y no tienen ningún efecto entre ellas. Entonces, una vez que las funciones de masa de probabilidad o de densidad de probabilidad marginales han sido especificadas, la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta es simplemente el producto de las dos funciones marginales. Se deduce que

$$P(a \le X \le b, c \le Y \le d) = P(a \le X \le b) \cdot P(c \le Y \le d)$$

Ejemplo 5.8 Suponga que las duraciones de dos componentes son independientes entre sí y que la distribución exponencial de la primera duración es X_1 con parámetro λ_1, mientras que la distribución exponencial de la segunda es X_2 con parámetro λ_2. Entonces la función de densidad de probabilidad conjunta es

$$\begin{aligned} f(x_1, x_2) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \\ &= \begin{cases} \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2} = \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{de lo contrario} \end{cases} \end{aligned}$$

Sean $\lambda_1 = 1/1000$ y $\lambda_2 = 1/1200$, de modo que las duraciones esperadas son 1000 y 1200 horas, respectivamente. La probabilidad de que ambas duraciones sean de por lo menos 1500 horas es

$$\begin{aligned} P(1500 \le X_1, 1500 \le X_2) &= P(1500 \le X_1) \cdot P(1500 \le X_2) \\ &= e^{-\lambda_1(1500)} \cdot e^{-\lambda_2(1500)} \\ &= (.2231)(.2865) = .0639 \end{aligned}$$

■

Más de dos variables aleatorias

Para modelar el comportamiento conjunto de más de dos variables aleatorias, se amplía el concepto de una distribución conjunta de dos variables.

DEFINICIÓN

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias discretas, la función de masa de probabilidad conjunta de las variables es la función

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Si las variables son continuas, la función de densidad de probabilidad conjunta de $X_1, \ldots, X_n$ es la función $f(x_1, x_2, \ldots, x_n)$ tal que para n intervalos cualesquiera $[a_1, b_1]$, $\ldots$, $[a_n, b_n]$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

En un experimento binomial, cada ensayo podría dar por resultado uno de sólo dos posibles resultados. Considérese ahora un experimento compuesto de n ensayos independientes e idénticos, en los que cada ensayo puede dar uno cualquiera de r posibles resultados. Sea $p_1 = P$(resultado i en cualquier ensayo particular) y defínanse las variables aleatorias como X_i = el número de ensayos que dan el resultado i ($i = 1, \ldots, r$). Tal experimento se llama **experimento multinomial** y la función de masa de probabilidad conjunta de $X_1, \ldots, X_r$ se llama **distribución multinomial.** Usando un argumento de conteo análogo al utilizado al deducir la distribución binomial, la función de masa de probabilidad conjunta de $X_1, \ldots, X_r$ se puede demostrar que es

$$p(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)(x_2!) \cdot \cdots \cdot (x_r!)} p_1^{x_1} \cdot \ldots \cdot p_r^{x_r} & x_i = 0, 1, 2, \ldots, \text{con } x_1 + \cdots + x_r = n \\ 0 & \text{de lo contrario} \end{cases}$$

El caso $r = 2$ da como resultado la distribución binomial, con X_1 = número de éxitos y $X_2 = n - X_1$ = número de fallas.

Ejemplo 5.9 Si se determina el alelo de cada una de diez secciones de un chícharo obtenidas independientemente y $p_1 = P(\text{AA})$, $p_2 = P(\text{Aa})$, $p_3 = P(\text{aa})$, $X_1 =$ número de AA, $X_2 =$ número de Aa y $X_3 =$ número de aa, entonces la función de masa de probabilidad multinomial para estas X_i es

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)(x_3!)} p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad x_i = 0, 1, \ldots \text{ y } x_1 + x_2 + x_3 = 10$$

Con $p_1 = p_3 = .25$, $p_2 = .5$.

$$\begin{aligned} P(X_1 = 2, X_2 = 5, X_3 = 3) &= p(2, 5, 3) \\ &= \frac{10!}{2!\,5!\,3!}(.25)^2(.5)^5(.25)^3 = .0769 \end{aligned}$$

■

Ejemplo 5.10 Cuando se utiliza cierto método para recolectar un volumen fijo de muestras de roca en una región, existen cuatro tipos de roca. Sean X_1, X_2 y X_3 la proporción por volumen de los tipos de roca 1, 2 y 3 en una muestra aleatoriamente seleccionada (la proporción del tipo de roca 4 es $1 - X_1 - X_2 - X_3$, de modo que una variable X_4 sería redundante). Si la función de densidad de probabilidad conjunta de X_1, X_2, X_3 es

$$f(x_1, x_2, x_3) = \begin{cases} kx_1x_2(1 - x_3) & 0 \le x_1 \le 1,\ 0 \le x_2 \le 1,\ 0 \le x_3 \le 1,\ x_1 + x_2 + x_3 \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

entonces k se determina como sigue

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1 \\ &= \int_0^1 \left\{ \int_0^{1-x_1} \left[\int_0^{1-x_1-x_2} kx_1x_2(1 - x_3)\, dx_3 \right] dx_2 \right\} dx_1 \end{aligned}$$

El valor de la integral iterada es $k/144$, por lo tanto $k = 144$. La probabilidad de que las rocas de los tipos 1 y 2 integren más de 50% de la muestra es

$$\begin{aligned} P(X_1 + X_2 \le .5) &= \iiint\limits_{\left\{ \substack{0 \le x_i \le 1 \text{ para } i=1,2,3 \\ x_1+x_2+x_3 \le 1,\ x_1+x_2 \le .5} \right\}} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1 \\ &= \int_0^{.5} \left\{ \int_0^{.5-x_1} \left[\int_0^{1-x_1-x_2} 144x_1x_2(1 - x_3)\, dx_3 \right] dx_2 \right\} dx_1 \\ &= .6066 \end{aligned}$$

■

La noción de independencia de más de dos variables aleatorias es similar a la noción de independencia de más de dos eventos.

DEFINICIÓN

Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son **independientes** si para *cada* subconjunto $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ de las variables (cada par, cada terna, y así sucesivamente), la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta del subconjunto es igual al producto de las funciones de masa de probabilidad o las funciones de densidad de probabilidad marginales.

Así pues, si las variables son independientes con $n = 4$, entonces la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta de dos variables cualesquiera es el producto de las dos marginales, y asimismo para tres variables cualesquiera y cuatro variables juntas. Aún más importante, una vez que se dice que n variables son independientes, entonces la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta es el producto de las n marginales.

Ejemplo 5.11 Si $X_1, \ldots, X_n$ representan las duraciones de n componentes y éstos operan de manera independiente uno de otro y cada duración está exponencialmente distribuida con parámetro λ, entonces para $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$,

$$f(x_1, x_2, \ldots, x_n) = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \Sigma x_i}$$

Si estos n componentes constituyen un sistema que fallará en cuanto un solo componente deje de funcionar, entonces la probabilidad de que el sistema dure más allá del tiempo t es

$$\begin{aligned} P(X_1 > t, \ldots, X_n > t) &= \int_t^\infty \ldots \int_t^\infty f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\ &= \left(\int_t^\infty \lambda e^{-\lambda x_1}\, dx_1 \right) \ldots \left(\int_t^\infty \lambda e^{-\lambda x_n}\, dx_n \right) \\ &= (e^{-\lambda t})^n = e^{-n\lambda t} \end{aligned}$$

Por consiguiente,

$$P(\text{duración del sistema } \leq t) = 1 - e^{-n\lambda t} \qquad \text{con } t \geq 0$$

lo que demuestra que la distribución de la duración del *sistema* es exponencial con parámetro $n\lambda$; el valor esperado de la duración del sistema es $1/n\lambda$. ■

En muchas situaciones experimentales que se considerarán en este libro, la independencia es una suposición razonable, de modo que la especificación de la distribución conjunta se reduce a decidir sobre distribuciones marginales apropiadas.

Distribuciones condicionales

Suponga $X =$ el número de defectos mayores en un automóvil nuevo seleccionado al azar y $Y =$ el número de defectos menores en el mismo auto. Si se sabe que el carro seleccionado tiene un defecto mayor, ¿cuál es ahora la probabilidad de que el carro tenga cuando mucho tres defectos menores?; es decir, ¿cuál es $P(Y \leq 3 \mid X = 1$? Asimismo, si X y Y denotan las duraciones de los neumáticos delantero y trasero de una motocicleta y sucede que $X = 10{,}000$ millas, ¿cuál es ahora la probabilidad de que Y sea cuando mucho de 15,000 millas y cuál es la duración esperada del neumático trasero "condicionada en" este valor de X? Preguntas de esta clase pueden ser respondidas estudiando distribuciones de probabilidad condicional.

DEFINICIÓN

Sean X y Y dos variables aleatorias continuas con función de densidad de probabilidad conjunta $f(x, y)$ y función de densidad de probabilidad marginal X, $f_X(x)$. Entonces para cualquier valor x de X para el cual $f_X(x) > 0$, la **función de densidad de probabilidad condicional de Y dado que $X = x$** es

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \qquad -\infty < y < \infty$$

Si X y Y son discretas, al reemplazar las funciones de densidad de probabilidad por funciones de masa de probabilidad en esta definición se obtiene la **función de masa de probabilidad condicional de Y cuando $X = x$.**

Obsérvese que la definición de $f_{Y|X}(y \mid x)$ es igual a la de $P(B \mid A)$, la probabilidad condicional de que B ocurra, dado que A ha ocurrido. Una vez que la función de densidad de probabilidad o la función de masa de probabilidad ha sido determinada, preguntas del tipo planteado al principio de esta subsección pueden ser respondidas integrando o sumando a lo largo de un conjunto apropiado de valores Y.

Ejemplo 5.12 Reconsidere la situación de los ejemplos 5.3 y 5.4 que implican X = la proporción del tiempo que la ventanilla para automovilista de un banco está ocupada y Y = la proporción análoga de la ventanilla normal. La función de densidad de probabilidad condicional de Y dado que $X = .8$ es

$$f_{Y|X}(y\,|\,.8) = \frac{f(.8, y)}{f_X(.8)} = \frac{1.2(.8 + y^2)}{1.2(.8) + .4} = \frac{1}{34}(24 + 30y^2) \quad 0 < y < 1$$

La probabilidad de que la ventanilla normal esté ocupada cuando mucho la mitad del tiempo dado que $X = .8$ es entonces

$$P(Y \le .5\,|\,X = .8) = \int_{-\infty}^{.5} f_{Y|X}(y\,|\,.8)\,dy = \int_0^{.5} \frac{1}{34}(24 + 30y^2)\,dy = .390$$

Utilizando la función de densidad de probabilidad marginal de Y se obtiene $P(Y \le .5) = .350$. Además $E(Y) = .6$, mientras que la proporción esperada del tiempo que la ventanilla normal está ocupada dado que $X = .8$ (una expectativa *condicional*) es

$$E(Y\,|\,X = .8) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y\,|\,.8)\,dy = \frac{1}{34}\int_0^1 y(24 + 30y^2)\,dy = .574$$ ■

Si las dos variables son independientes, las funciones de masa de probabilidad o de densidad de probabilidad marginales en el denominador, cancelarán el factor correspondiente en el numerador. La distribución condicional es entonces idéntica a la distribución marginal correspondiente.

EJERCICIOS Sección 5.1 (1–21)

1. Una gasolinera cuenta tanto con islas de autoservicio como de servicio completo. En cada isla hay una sola bomba de gasolina sin plomo regular con dos mangueras. Sea X el número de mangueras utilizadas en la isla de autoservicio en un tiempo particular y sea Y el número de mangueras en uso en la isla de servicio completo en ese tiempo. La función de masa de probabilidad conjunta de X y Y aparece en la tabla adjunta.

$p(x, y)$		$y = 0$	$y = 1$	$y = 2$
	0	.10	.04	.02
x	1	.08	.20	.06
	2	.06	.14	.30

a. ¿Cuál es $P(X = 1 \text{ y } Y = 1)$?
b. Calcule $P(X \le 1 \text{ y } Y \le 1)$.
c. Describa con palabras el evento ($X \ne 0$ y $Y \ne 0$) y calcule su probabilidad.
d. Calcule la función de masa de probabilidad marginal de X y Y. Utilizando $p_X(x)$, ¿cuál es $P(X \le 1)$?
e. ¿Son X y Y variables aleatorias independientes? Explique.

2. Cuando un automóvil es detenido por una patrulla de seguridad, cada uno de los neumáticos es revisado en cuanto a desgaste y cada uno de los faros es revisado para ver si está apropiadamente alineado. Sean X el número de faros que necesitan ajuste y Y el número de neumáticos defectuosos.
a. Si X y Y son independientes con $p_X(0) = .5$, $p_X(1) = .3$, $p_X(2) = .2$ y $p_Y(0) = .6$, $p_Y(1) = .1$, $p_Y(2) = p_Y(3) = .05$ y $p_Y(4) = .2$, muestre la función de masa de probabilidad conjunta de (X, Y) en una tabla de probabilidad conjunta.
b. Calcule $P(X \le 1 \text{ y } Y \le 1)$ con la tabla de probabilidad conjunta y verifique que es igual al producto $P(X \le 1) \cdot P(Y \le 1)$.
c. ¿Cuál es $P(X + Y = 0)$ (la probabilidad de ninguna violación)?
d. Calcule $P(X + Y \le 1)$.

3. Un supermercado cuenta tanto con una caja rápida como con una extrarrápida. Sea X_1 el número de clientes formados en la caja rápida a una hora particular del día y sea X_2 el número de clientes formados en la caja extrarrápida a la misma hora. Suponga que la función de masa de probabilidad conjunta de X_1 y X_2 es la que aparece en la tabla adjunta.

		$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$
	0	.08	.07	.04	.00
	1	.06	.15	.05	.04
x_1	2	.05	.04	.10	.06
	3	.00	.03	.04	.07
	4	.00	.01	.05	.06

a. ¿Cuál es $P(X_1 = 1, X_2 = 1)$; es decir, la probabilidad de que haya exactamente un cliente en cada caja?
b. ¿Cuál es $P(X_1 = X_2)$; es decir, la probabilidad de que los números de clientes en las dos cajas sean idénticos?
c. Sea A el evento en que hay por lo menos dos clientes más en una caja que en la otra. Exprese A en función de X_1 y X_2, y calcule la probabilidad de este evento.
d. ¿Cuál es la probabilidad de que el número total de clientes en las dos líneas sea exactamente cuatro? ¿Por lo menos cuatro?

4. Regrese a la situación descrita en el ejercicio 3.

a. Determine la función de masa de probabilidad marginal de X_1 y en seguida calcule el número esperado de clientes formados en la caja rápida.

b. Determine la función de masa de probabilidad marginal de X_2.

c. Por inspección de las probabilidades $P(X_1 = 4)$, $P(X_2 = 0)$ y $P(X_1 = 4, X_2 = 0)$, ¿son X_1 y X_2 variables aleatorias independientes? Explique.

5. El número de clientes que esperan en el servicio de envoltura de regalos en una tienda de departamentos es una variable aleatoria X con valores posibles 0, 1, 2, 3, 4 y probabilidades correspondientes .1, .2, .3, .25, .15. Un cliente seleccionado al azar tendrá 1, 2 o 3 paquetes para envoltura con probabilidades de .6, .3 y .1, respectivamente. Sea Y = el número total de paquetes que van a ser envueltos para los clientes que esperan formados en la fila (suponga que el número de paquetes entregado por un cliente es independiente del número entregado por cualquier otro cliente).

a. Determine $P(X = 3, Y = 3)$, es decir, $p(3, 3)$.

b. Determine $p(4, 11)$.

6. Sea X el número de cámaras digitales Canon vendidas durante una semana particular por una tienda. La función de masa de probabilidad de X es

x	0	1	2	3	4
$p_X(x)$	.1	.2	.3	.25	.15

60% de todos los clientes que compran estas cámaras también compran una garantía extendida. Sea Y el número de compradores durante esta semana que compran una garantía extendida.

a. ¿Cuál es $P(X = 4, Y = 2)$? [*Sugerencia*: esta probabilidad es igual a $P(Y = 2 \mid X = 4) \cdot P(X = 4)$; ahora piense en las cuatro compras como cuatro ensayos de un experimento binomial, con el éxito en un ensayo correspondiente a comprar una garantía extendida.]

b. Calcule $P(X = Y)$.

c. Determine la función de masa de probabilidad conjunta de X y Y y luego la función de masa de probabilidad marginal de Y.

7. La distribución de probabilidad conjunta del número X de carros y el número Y de autobuses por ciclo de señal en un carril de vuelta a la izquierda propuesto se muestra en la tabla de probabilidad conjunta anexa.

$p(x, y)$		y = 0	1	2
	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
x	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

a. ¿Cuál es la probabilidad de que haya exactamente un carro y exactamente un autobús durante un ciclo?

b. ¿Cuál es la probabilidad de que haya cuando mucho un carro y cuando mucho un autobús durante un ciclo?

c. ¿Cuál es la probabilidad de que haya exactamente un carro durante un ciclo? ¿Exactamente un autobús?

d. Suponga que el carril de vuelta a la izquierda tiene una capacidad de cinco carros y que un autobús equivale a tres carros. ¿Cuál es la probabilidad de un exceso de vehículos durante un ciclo?

e. ¿Son X y Y variables aleatorias independientes? Explique.

8. Un almacén cuenta con 30 componentes de un tipo, de los cuales 8 fueron surtidos por el proveedor 1, 10 por el proveedor 2 y 12 por el proveedor 3. Seis de éstos tienen que ser seleccionados al azar para un ensamble particular. Sea X = el número de componentes del proveedor 1 seleccionados, Y = el número de componentes del proveedor 2 seleccionados y que $p(x, y)$ denote la función de masa de probabilidad conjunta de X y Y.

a. ¿Cuál es $p(3, 2)$? [*Sugerencia*: la probabilidad de que cada muestra de tamaño 6 sea seleccionada es igual. Por consiguiente, $p(3, 2)$ = (número de resultados con $X = 3$ y $Y = 2$)/(el número total de resultados). Ahora use la regla de producto para el conteo para obtener el numerador y el denominador.]

b. Utilizando la lógica del inciso (a), obtenga $p(x, y)$. (Esto puede ser considerado como un muestreo con distribución hipergeométrica multivariante sin reemplazo de una población finita compuesta por más de dos categorías.)

9. Se supone que cada neumático delantero de un tipo particular de vehículo está inflado a una presión de 26 lb/pulg². Suponga que la presión de aire real en cada neumático es una variable aleatoria, X para el neumático derecho y Y para el izquierdo con función de densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{de lo contrario} \end{cases}$$

a. ¿Cuál es el valor de K?

b. ¿Cuál es la probabilidad de que ambos neumáticos estén inflados a menos presión?

c. ¿Cuál es la probabilidad de que la diferencia en la presión del aire entre los dos neumáticos sea de cuando mucho 2 lb/pulg²?

d. Determine la distribución (marginal) de la presión del aire sólo en el neumático derecho.

e. ¿Son X y Y variables aleatorias independientes?

10. Annie y Alvie acordaron encontrarse entre las 5:00 p.m. y las 6:00 p.m. para cenar en un restaurante local de comida saludable. Sea X = la hora de llegada de Annie y Y = la hora de llegada de Alvie. Suponga que X y Y son independientes, cada una distribuida uniformemente en el intervalo [5, 6].

a. ¿Cuál es la función de densidad de probabilidad conjunta de X y Y?

b. ¿Cuál es la probabilidad de que ambas lleguen entre las 5:15 y las 5:45?

c. Si la primera en llegar espera sólo 10 min antes de irse a comer a otra parte, ¿cuál es la probabilidad de que cenen en el restaurante de comida saludable? [*Sugerencia*: el evento de interés es $A = \{(x, y): |x - y| \le \frac{1}{6}\}$.]

11. Dos profesores acaban de entregar los exámenes finales para su copia. Sea X el número de errores tipográficos en el examen del primer profesor y Y el número de tales errores en el segundo examen. Suponga que X tiene una distribución de Poisson con parámetro μ_1, que Y tiene una distribución de Poisson con parámetro μ_2 y que X y Y son independientes.

a. ¿Cuál es la función de masa de probabilidad conjunta de X y Y?

b. ¿Cuál es la probabilidad de que se cometa cuando mucho un error en ambos exámenes combinados?

c. Obtenga una expresión general para la probabilidad de que el número total de errores en los dos exámenes sea m (donde m es un entero no negativo). [*Sugerencia*: $A = \{(x, y): x + y = m\} = \{(m, 0), (m - 1, 1), \ldots, (1, m - 1), (0, m)\}$. Ahora sume la función de masa de probabilidad conjunta a lo largo de $(x, y) \in A$ y use el teorema binomial, el cual dice que

$$\sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k} = (a + b)^m$$

para cualquier a, b.]

12. Dos componentes de una minicomputadora tienen la siguiente función de densidad de probabilidad conjunta de sus vidas útiles X y Y:

$$f(x, y) = \begin{cases} xe^{-x(1+y)} & x \geq 0 \text{ y } y \geq 0 \\ 0 & \text{de lo contrario} \end{cases}$$

a. ¿Cuál es la probabilidad de que la vida útil X del primer componente exceda de 3?

b. ¿Cuáles son las funciones de densidad de probabilidad marginales de X y Y? ¿Son las dos vidas útiles independientes? Explique.

c. ¿Cuál es la probabilidad de que la vida útil de por lo menos un componente exceda de 3?

13. Tiene dos focos para una lámpara particular. Sea X = la vida útil del primer foco y Y = la vida útil del segundo (ambas en miles de horas). Suponga que X y Y son independientes y que cada una tiene una distribución exponencial con parámetro $\lambda = 1$.

a. ¿Cuál es la función de densidad de probabilidad conjunta de X y Y?

b. ¿Cuál es la probabilidad de que cada foco dure cuando mucho 1000 horas (es decir, $X \leq 1$ y $Y \leq 1$)?

c. ¿Cuál es la probabilidad de que la vida útil total de los dos focos sea cuando mucho de 2? [*Sugerencia*: trace una figura de la región $A = \{(x, y): x \geq 0, y \geq 0, x + y \leq 2\}$ antes de integrar.]

d. ¿Cuál es la probabilidad de que la vida útil total sea de entre 1 y 2?

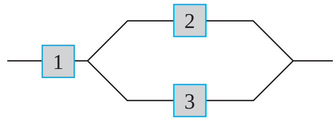
14. Suponga que tiene diez focos y que la vida útil de cada uno es independiente de la de los demás y que cada vida útil tiene una distribución exponencial con parámetro λ.

a. ¿Cuál es la probabilidad de que los diez focos fallen antes del tiempo t?

b. ¿Cuál es la probabilidad de que exactamente k de los diez focos fallen antes del tiempo t?

c. Suponga que nueve de los focos tienen vidas útiles exponencialmente distribuidas con parámetro λ y que el foco restante tiene una vida útil que está exponencialmente distribuida con parámetro θ (fue hecho por otro fabricante). ¿Cuál es la probabilidad de que exactamente cinco de los diez focos fallen antes del tiempo t?

15. Considere un sistema compuesto de tres componentes como se ilustra. El sistema continuará funcionando en tanto el primer componente funcione y el componente 2 o el componente 3 funcionen. Sean X_1, X_2 y X_3 las vidas útiles de los componentes 1, 2 y 3, respectivamente. Suponga que las X_i son independientes una de otra y que cada X_i tiene una distribución exponencial con parámetro λ.



a. Sea Y la vida útil del sistema. Obtenga la función de distribución acumulativa de Y y diferénciela para obtener la función de densidad de probabilidad. [*Sugerencia*: $F(y) = P(Y \leq y)$; exprese el evento $\{Y \leq y\}$ en función de uniones y/o intersecciones de los tres eventos $\{X_1 \leq y\}$, $\{X_2 \leq y\}$ y $\{X_3 \leq y\}$.]

b. Calcule la vida útil esperada del sistema.

16. **a.** Para $f(x_1, x_2, x_3)$ del ejemplo 5.10, calcule la **función de densidad marginal conjunta** sólo de X_1 y X_3 (integrando para x_2).

b. ¿Cuál es la probabilidad de que rocas de los tipos 1 y 3 constituyan cuando mucho 50% de la muestra? [*Sugerencia*: use el resultado del inciso (a).]

c. Calcule la función de densidad de probabilidad marginal sólo de X_1 [*Sugerencia*: use el resultado del inciso (a).]

17. Un ecólogo desea seleccionar un punto dentro de una región de muestreo circular de acuerdo con una distribución uniforme (en la práctica esto podría hacerse seleccionando primero una dirección y luego una distancia a partir del centro en esa dirección). Sea X = la coordenada x del punto seleccionado y Y = la coordenada y del punto seleccionado. Si el círculo tiene su centro en (0, 0) y su radio es R, entonces la función de densidad de probabilidad conjunta de X y Y es

$$f(x, y) = \begin{cases} \dfrac{1}{\pi R^2} & x^2 + y^2 \leq R^2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. ¿Cuál es la probabilidad de que el punto seleccionado quede dentro de $R/2$ del centro de la región circular? [*Sugerencia*: trace una figura de la región de densidad positiva D. Como $f(x, y)$ es constante en D, el cálculo de probabilidad se reduce al cálculo de un área.]

b. ¿Cuál es la probabilidad de que tanto X como Y difieran de 0 por cuando mucho $R/2$?

c. Responda el inciso (b) con $R/\sqrt{2}$ reemplazando a $R/2$.

d. ¿Cuál es la función de densidad de probabilidad marginal de X? ¿De Y? ¿Son X y Y independientes?

18. Remítase al ejercicio 1 y responda las siguientes preguntas:

a. Dado que $X = 1$, determine la función de masa de probabilidad condicional de Y; es decir, $p_{Y|X}(0 \mid 1)$, $p_{Y|X}(1 \mid 1)$ y $p_{Y|X}(2 \mid 1)$.

b. Dado que dos mangueras están en uso en la isla de autoservicio, ¿cuál es la función de masa de probabilidad condicional del número de mangueras en uso en la isla de servicio completo?

c. Use el resultado del inciso (b) para calcular la probabilidad condicional $P(Y \leq 1 \mid X = 2)$.

d. Dado que dos mangueras están en uso en la isla de servicio completo, ¿cuál es la función de masa de probabilidad condicional del número en uso en la isla de autoservicio?

19. La función de densidad de probabilidad conjunta de las presiones de los neumáticos delanteros derecho e izquierdo se da en el ejercicio 9.

a. Determine la función de densidad de probabilidad condicional de Y dado que $X = x$ y la función de densidad de probabilidad condicional de X dado que $Y = y$.

b. Si la presión del neumático derecho es de 22 lb/pulg^2, ¿cuál es la probabilidad de que la presión del neumático izquierdo sea de por lo menos 25 lb/pulg^2? Compare esto con $P(Y \geq 25)$.

c. Si la presión del neumático derecho es de 22 lb/pulg^2, ¿cuál es la presión esperada en el neumático izquierdo y cuál es la desviación estándar de la presión en este neumático?

20. Sean X_1, X_2, X_3, X_4, X_5 y X_6 los números de caramelos M&M azules, cafés, verdes, naranjas, rojos y amarillos, respectivamente, en una muestra de tamaño n. Entonces estas X_i tienen una distribución multinomial. De acuerdo con el sitio web de M&M, las proporciones de colores son $p_1 = .24$, $p_2 = .13$, $p_3 = .16$, $p_4 = .20$, $p_5 = .13$ y $p_6 = .14$.

a. Si $n = 12$, ¿cuál es la probabilidad de que haya exactamente dos caramelos M&M de cada color?

b. Con $n = 20$, ¿cuál es la probabilidad de que haya cuando mucho cinco caramelos naranja? [*Sugerencia*: considere el caramelo naranja como un éxito y cualquier otro color como falla.]

c. En una muestra de 20 caramelos M&M, ¿cuál es la probabilidad de que el número de caramelos azules, verdes o naranjas sea por lo menos de 10?

21. Sean X_1, X_2 y X_3 las vidas útiles de los componentes 1, 2 y 3 en un sistema de tres componentes.

a. ¿Cómo definiría la función de densidad de probabilidad condicional de X_3 dado que $X_1 = x_1$ y $X_2 = x_2$?

b. ¿Cómo definiría la función de densidad de probabilidad conjunta condicional de X_2 y X_3 dado que $X_1 = x_1$?

5.2 Valores esperados, covarianza y correlación

Previamente se vio que cualquier función $h(x)$ de una sola variable aleatoria X es por sí misma una variable aleatoria. Sin embargo, para calcular $E[h(X)]$, no es necesario obtener la distribución de probabilidad de $h(X)$; en cambio, $E[h(X)]$ se calculó como un promedio ponderado de valores de $h(x)$, donde la función de ponderación fue la función de masa de probabilidad $p(x)$ o la función de densidad de probabilidad $f(x)$ de X. Se obtiene un resultado similar para una función $h(X, Y)$ de dos variables aleatorias conjuntamente distribuidas.

PROPOSICIÓN

Sean X y Y variables aleatorias conjuntamente distribuidas con función de masa de probabilidad $p(x, y)$ o función de densidad de probabilidad $f(x, y)$ ya sea que las variables sean discretas o continuas. Entonces el valor esperado de una función $h(X, Y)$ denotada por $E[h(X, Y)]$ o $\mu_{h(X,Y)}$ está dada por

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & \text{si } X \text{ y } Y \text{ son discretas} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{si } X \text{ y } Y \text{ son continuas} \end{cases}$$

Ejemplo 5.13 Cinco amigos compraron boletos para un concierto. Si los boletos son para los asientos 1–5 en una fila particular y los boletos se distribuyen al azar entre los cinco, ¿cuál es el número esperado de asientos que separen a cualesquiera dos de los cinco? Sean X y Y los números de asiento del primero y segundo individuos, respectivamente. Los pares posibles (X, Y) son $\{(1, 2), (1, 3), \ldots, (5, 4)\}$ y la función de masa de probabilidad conjunta de (X, Y) es

$$p(x, y) = \begin{cases} \dfrac{1}{20} & x = 1, \ldots, 5;\ y = 1, \ldots, 5;\ x \neq y \\ 0 & \text{de lo contrario} \end{cases}$$

El número de asientos que separan a los dos individuos es $h(X, Y) = |X - Y| - 1$. La tabla adjunta da $h(x, y)$ para cada posible par (x, y).

$h(x, y)$		x = 1	2	3	4	5
	1	—	0	1	2	3
	2	0	—	0	1	2
y	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

Por lo tanto

$$E[h(X, Y)] = \sum_{(x,y)}\sum h(x, y) \cdot p(x, y) = \sum_{\substack{x=1 \\ x \neq y}}^{5} \sum_{y=1}^{5} (|x - y| - 1) \cdot \frac{1}{20} = 1 \quad ■$$

Ejemplo 5.14 En el ejemplo 5.5, la función de densidad de probabilidad conjunta de la cantidad X de almendras y la cantidad Y de nueces de acajú en una lata de 1 lb de nueces fue

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Si 1 lb de almendras le cuesta a la compañía \$1.00, 1 lb de nuez de acajú le cuesta \$1.50 y 1 lb de cacahuates le cuesta \$.50, entonces el costo total del contenido de una lata es

$$h(X, Y) = (1)X + (1.5)Y + (.5)(1 - X - Y) = .5 + .5X + Y$$

(puesto que $1 - X - Y$ del peso se compone de cacahuates). El costo esperado total es

$$\begin{aligned} E[h(X, Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\,dx\,dy \\ &= \int_0^1\int_0^{1-x} (.5 + .5x + y) \cdot 24xy\,dy\,dx = \$1.10 \quad ■ \end{aligned}$$

El método de calcular el valor esperado de una función $h(X_1, \ldots, X_n)$ de n variables aleatorias es similar al de dos variables aleatorias. Si las X_i son discretas, $E[h(X_1, \ldots, X_n)]$ es una suma de n dimensiones; si las X_i son continuas, es una integral de n dimensiones.

Covarianza

Cuando dos variables aleatorias X y Y no son independientes, con frecuencia es de interés valorar qué tan fuerte están relacionadas una con otra.

DEFINICIÓN

La **covarianza** entre dos variables aleatorias X y Y es

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ discretas} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\,dx\,dy & X, Y \text{ continuas} \end{cases}$$

Es decir, como $X - \mu_X$ y $Y - \mu_Y$ son las desviaciones de las dos variables con respecto a sus valores medios, la covarianza es el producto esperado de las desviaciones. Obsérvese que $\text{Cov}(X, X) = E[(X - \mu_X)^2] = V(X)$.

La exposición razonada para la definición es como sigue. Suponga que X y Y tienen una fuerte relación positiva entre ellas, lo que significa que los valores grandes de X tienden a ocurrir con valores grandes de Y y los valores pequeños de X con los valores pequeños de Y. Entonces la mayor parte de la masa o densidad de probabilidad estará asociada para $(x - \mu_X)$ y $(y - \mu_Y)$, ambos positivos (tanto X como Y por arriba de sus respectivas medias) o ambos negativos, así que el producto $(x - \mu_X)(y - \mu_Y)$ tenderá a ser positivo. Por tanto para una fuerte relación positiva, $\text{Cov}(X, Y)$ deberá ser bastante positiva. Con una fuerte relación negativa los signos de $(x - \mu_X)$ y $(y - \mu_Y)$ tenderán a ser opuestos, lo que da un producto negativo. Por tanto, con una fuerte relación negativa, $\text{Cov}(X, Y)$ deberá ser bastante negativa. Si X y Y no están fuertemente relacionadas, los productos positivo y negativo tenderán a eliminarse entre sí, lo que da una covarianza de cerca de 0. La figura 5.4 ilustra las diferentes posibilidades. La covarianza depende *tanto* del conjunto de pares posibles *como* de las probabilidades. En la figura 5.4, las probabilidades podrían ser cambiadas sin que se altere el conjunto de pares posibles y esto podría cambiar drásticamente el valor de $\text{Cov}(X, Y)$.

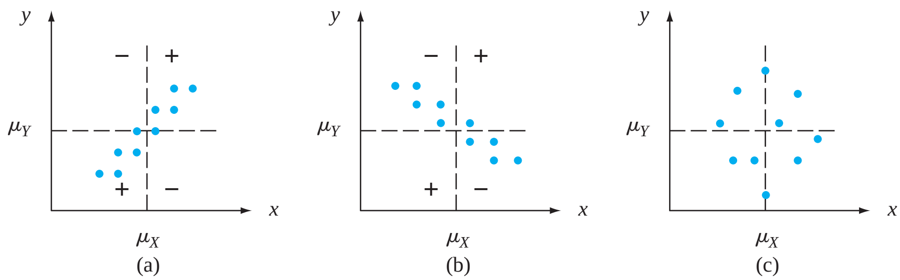


Figura 5.4 $p(x, y) = 1/10$ de cada uno de los diez pares correspondientes a los puntos indicados; (a) covarianza positiva; (b) covarianza negativa; (c) covarianza cerca de cero

Ejemplo 5.15 Las funciones de masa de probabilidad conjunta y marginal de X = cantidad deducible sobre una póliza de automóvil y Y = cantidad deducible sobre una póliza de propietario de casa en el ejemplo 5.1 fueron

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	200
$p_Y(y)$	.25	.25	.5

de las cuales $\mu_X = \Sigma x p_X(x) = 175$ y $\mu_Y = 125$. Por consiguiente,

$$\begin{aligned}\text{Cov}(X, Y) &= \sum_{(x,y)}\sum (x - 175)(y - 125)p(x, y) \\ &= (100 - 175)(0 - 125)(.20) + \cdots \\ &\quad + (250 - 175)(200 - 125)(.30) \\ &= 1875\end{aligned}$$

■

La siguiente fórmula abreviada para $\text{Cov}(X, Y)$ simplifica los cálculos.

PROPOSICIÓN

$$\text{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

De acuerdo con esta fórmula, no se requieren sustracciones intermedias; sólo al final del cálculo se resta $\mu_X \cdot \mu_Y$ de $E(XY)$. La comprobación implica expandir $(X - \mu_X)(Y - \mu_Y)$ y luego considerar el valor esperado de cada término por separado.

Ejemplo 5.16 (Continuación del ejemplo 5.5)

Las funciones de densidad de probabilidad conjunta y marginal de X = cantidad de almendras y Y = cantidad de nueces de acajú fueron

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1 - x)^2 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

con $f_Y(y)$ obtenida reemplazando x por y en $f_X(x)$. Es fácil verificar que $\mu_X = \mu_Y = \frac{2}{5}$ y

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy = \int_0^1 \int_0^{1-x} xy \cdot 24xy\, dy\, dx$$

$$= 8 \int_0^1 x^2(1 - x)^3\, dx = \frac{2}{15}$$

Por lo tanto $\text{Cov}(X, Y) = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}$. Una covarianza negativa se considera razonable en este caso porque más almendras contenidas en la lata implican menos nueces de acajú. ■

Pudiera parecer que la relación en el ejemplo de los seguros es bastante fuerte puesto que $\text{Cov}(X, Y) = 1875$, mientras que $\text{Cov}(X, Y) = -\frac{2}{75}$ en el ejemplo de las nueces parecería implicar una relación bastante débil. Desafortunadamente, la covarianza tiene un serio defecto que hace imposible interpretar un valor calculado. En el ejemplo de los seguros, suponga que la cantidad deducible se expresó en centavos en lugar de en dólares. Entonces $100X$ reemplazaría a X, $100Y$ reemplazaría a Y y la covarianza resultante sería $\text{Cov}(100X, 100Y) = (100)(100)\text{Cov}(X, Y) = 18{,}750{,}000$. Si, por otra parte, la cantidad deducible se hubiera expresado en cientos de dólares, la covarianza calculada habría sido $(.01)(.01)(1875) = .1875$. *El defecto de la covarianza es que su valor calculado depende críticamente de las unidades de medición*. De manera ideal, la selección de las unidades no debe tener efecto en la medida de la fuerza de la relación. Esto se logra graduando a escala la covarianza.

Correlación

DEFINICIÓN

El **coeficiente de correlación** de X y Y, denotado por $\text{Corr}(X, Y)$, $\rho_{X,Y}$, o simplemente ρ, está definido por

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Ejemplo 5.17

Es fácil verificar que en el escenario de los seguros del ejemplo 5.15, $E(X^2) = 36{,}250$, $\sigma_X^2 = 36{,}250 - (175)^2 = 5625$, $\sigma_X = 75$, $E(Y^2) = 22{,}500$, $\sigma_Y^2 = 6875$ y $\sigma_Y = 82.92$. Esto da

$$\rho = \frac{1875}{(75)(82.92)} = .301$$ ■

La siguiente proposición muestra que ρ remedia el defecto de Cov(X, Y) y también sugiere cómo reconocer la existencia de una fuerte relación (lineal).

PROPOSICIÓN

1. Si a y c son ambas positivas o ambas negativas,
$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$
2. Para dos variables aleatorias X y Y cualesquiera, $-1 \leq \text{Corr}(X, Y) \leq 1$.

La proposición 1 dice precisamente que el coeficiente de correlación no se ve afectado por un cambio lineal en las unidades de medición (si, por ejemplo, X = temperatura en °C, entonces $9X/5 + 32$ = temperatura en °F). De acuerdo con la proposición 2, la relación positiva más fuerte posible es puesta en evidencia por $\rho = +1$, en tanto que la relación negativa más fuerte posible corresponde a $\rho = -1$. La comprobación de la primera proposición se ilustra en el ejercicio 35, y la de la segunda aparece en el ejercicio suplementario 87 al final del capítulo. Para propósitos descriptivos, la relación se describirá como fuerte si $|\rho| \geq .8$, moderada si $.5 < |\rho| < .8$ y débil si $|\rho| \leq .5$.

Si se considera que $p(x, y)$ o $f(x, y)$ prescribe un modelo matemático de cómo las dos variables numéricas X y Y están distribuidas en alguna población (estatura y peso, calificaciones del examen de aptitud escolar verbal y cuantitativa, etc.), entonces ρ es una característica o parámetro de población que mide cuán fuertemente están relacionadas X y Y en la población. En el capítulo 12 se considerará tomar una muestra de pares $(x_1, y_1), \ldots, (x_n, y_n)$ de la población. El coeficiente de correlación muestral r se definirá y utilizará entonces para hacer inferencias con respecto a ρ.

El coeficiente de correlación ρ no es en realidad una medida completamente general de la fuerza de una relación.

PROPOSICIÓN

1. Si X y Y son independientes, entonces $\rho = 0$, pero $\rho = 0$ no implica independencia.
2. $\rho = 1$ o -1 si y sólo si $Y = aX + b$ con algunos números a y b con $a \neq 0$.

Esta proposición dice que ρ mide el grado de relación **lineal** entre X y Y, y sólo cuando las dos variables están perfectamente relacionadas de una manera lineal, ρ será tan positivo o negativo como pueda serlo. Un ρ menor que 1 en valor absoluto indica sólo que la relación no es completamente lineal, pero que aún puede haber una fuerte relación no lineal. Además, $\rho = 0$ no implica que X y Y sean independientes, sino sólo que existe una ausencia completa de relación lineal. Cuando $\rho = 0$, se dice que X y Y **no están correlacionadas.** Dos variables podrían no estar correlacionadas y no obstante ser altamente dependientes porque existe una fuerte relación no lineal, así que se debe tener cuidado de no concluir demasiado del hecho de que $\rho = 0$.

Ejemplo 5.18 Sean X y Y variables aleatorias discretas con función de masa de probabilidad conjunta

$$p(x, y) = \begin{cases} \frac{1}{4} & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{de lo contrario} \end{cases}$$

Los puntos que reciben masa de probabilidad positiva están identificados en el sistema de coordenadas (x, y) en la figura 5.5. Es evidente por la figura que el valor de X está completamente determinado por el valor de Y y viceversa, de modo que las dos variables son completamente dependientes. Sin embargo, por simetría $\mu_X = \mu_Y = 0$ y $E(XY) = (-4)\frac{1}{4} + (-4)\frac{1}{4} + (4)\frac{1}{4} + (4)\frac{1}{4} = 0$, la covarianza es por tanto $\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y = 0$ y por consiguiente $\rho_{X,Y} = 0$. ¡Aunque hay una dependencia perfecta, también hay una ausencia completa de cualquier relación lineal!

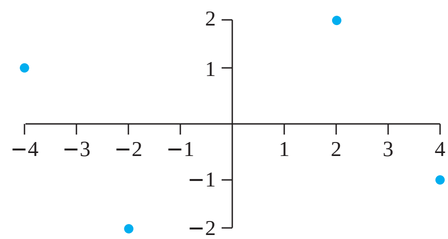


Figura 5.5 Población de pares del ejemplo 5.18 ■

Un valor de ρ próximo a 1 no necesariamente implica que el incremento del valor de X *hace* que se incremente Y. Implica sólo que los valores grandes de X están *asociados* con valores grandes de Y. Por ejemplo, en la población de niños, el tamaño del vocabulario y el número de caries no están lo bastante correlacionados positivamente, pero con certeza no es cierto que las caries hagan que crezca el vocabulario. En cambio, los valores de estas dos variables tienden a incrementarse conforme el valor de la edad, una tercera variable, se incrementa. Para niños de una edad fija, quizás existe una muy baja correlación entre el número de caries y el tamaño del vocabulario. En suma, asociación (una alta correlación) no es lo mismo que causa.

EJERCICIOS Sección 5.2 (22–36)

22. Un instructor aplicó un corto examen compuesto de dos partes. Para un estudiante seleccionado al azar, sea X = el número de puntos obtenidos en la primera parte y Y = el número de puntos obtenidos en la segunda parte. Suponga que la función de masa de probabilidad conjunta de X y Y se da en la tabla adjunta.

			y		
$p(x, y)$		0	5	10	15
	0	.02	.06	.02	.10
x	5	.04	.15	.20	.10
	10	.01	.15	.14	.01

a. Si la calificación anotada en la libreta de calificaciones es el número total de puntos obtenidos en las dos partes, ¿cuál es la calificación anotada esperada $E(X + Y)$?

b. Si se anota la máxima de las dos calificaciones, ¿cuál es la calificación anotada esperada?

23. La diferencia entre el número de clientes formados en la caja rápida y el número formado en la caja extrarrápida del ejercicio 3 es $X_1 - X_2$. Calcule la diferencia esperada.

24. Seis individuos, incluidos A y B, se sientan alrededor de una mesa circular en una forma completamente al azar. Suponga que los asientos están numerados 1, . . . , 6. Sea X = el número de asiento de A y Y = el número de asiento de B. Si A envía un mensaje escrito alrededor de la mesa a B en la dirección en la cual están más cerca, ¿cuántos individuos (incluidos A y B) esperaría que manipulen el mensaje?

25. Un topógrafo desea delimitar una región cuadrada con longitud de lado L. Sin embargo, debido a un error de medición, delimita en cambio un rectángulo en el cual los lados norte–sur son de longitud X y los lados este–oeste son de longitud Y. Suponga que X y Y son independientes y que cada uno está uniformemente distribuido en el intervalo $[L - A, L + A]$ (donde $0 < A < L$). ¿Cuál es el área esperada del rectángulo resultante?

26. Considere un pequeño transbordador que puede transportar carros y autobuses. La cuota para carros es de \$3 y para autobuses es de \$10. Sean X y Y el número de carros y autobuses, respectivamente, transportados en un solo viaje. Suponga que la distribución conjunta de X y Y aparece en la tabla del ejercicio 7. Calcule el ingreso esperado en un solo viaje.

27. Annie y Alvie quedaron de encontrarse para desayunar entre el mediodía (0:00 p.m.) y 1:00 p.m. Denote la hora de llegada de Annie por X, y la de Alvie por Y y suponga que X y Y son independientes con funciones de densidad de probabilidad

$$f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es la cantidad esperada de tiempo que la que llega primero debe esperar a la otra persona? [*Sugerencia*: $h(X, Y) = |X - Y|$.]

28. Demuestre que si X y Y son variables aleatorias independientes, entonces $E(XY) = E(X) \cdot E(Y)$. Luego aplique esto en el ejercicio 25. [*Sugerencia*: considere el caso continuo con $f(x, y) = f_X(x) \cdot f_Y(y)$.]

29. Calcule el coeficiente de correlación ρ de X y Y del ejemplo 5.16 (ya se calculó la covarianza).

30. **a.** Calcule la covarianza de X y Y en el ejercicio 22.
b. Calcule ρ para X y Y en el mismo ejercicio.

31. **a.** Calcule la covarianza entre X y Y en el ejercicio 9.
b. Calcule el coeficiente de correlación ρ para X y Y.

32. Reconsidere las vidas útiles de los componentes de minicomputadora X y Y como se describe en el ejercicio 12. Determine $E(XY)$. ¿Qué se puede decir sobre $\text{Cov}(X, Y)$ y ρ?

33. Use el resultado del ejercicio 28 para demostrar que cuando X y Y son independientes, $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$.

34. **a.** Recordando la definición de σ^2 para una sola variable aleatoria X, escriba una fórmula que sería apropiada para calcular la varianza de una función $h(X, Y)$ de dos variables aleatorias. [*Sugerencia*: recuerde que la varianza es simplemente un valor esperado especial.]
b. Use esta fórmula para calcular la varianza de la calificación anotada $h(X, Y)$ [$=$ máx(X, Y) en el inciso (b) del ejercicio 22.

35. **a.** Use las reglas de valor esperado para demostrar que $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$.
b. Use el inciso (a) junto con las reglas de varianza y desviación estándar para demostrar que $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$ cuando a y c tienen el mismo signo.
c. ¿Qué sucede si a y c tienen signos opuestos?

36. Demuestre que si $Y = aX + b$ $(a \ne 0)$, entonces $\text{Corr}(X, Y) = +1$ o -1. ¿En qué condiciones será $\rho = +1$?

5.3 Estadísticos y sus distribuciones

En el capítulo 1, $x_1, x_2, \ldots, x_n$ denotaron las observaciones en una sola muestra. Considérese seleccionar dos muestras diferentes de tamaño n de la misma distribución de población. Las x_i en la segunda muestra diferirán siempre virtualmente por lo menos un poco de aquellas en la primera muestra. Por ejemplo, una primera muestra de $n = 3$ carros de un tipo particular podría producir eficiencias de combustible $x_1 = 30.7$, $x_2 = 29.4$, $x_3 = 31.1$, mientras que una segunda muestra puede dar $x_1 = 28.8$, $x_2 = 30.0$ y $x_3 = 32.5$. Antes de obtener datos, existe incertidumbre sobre el valor de cada x_i. Debido a esta incertidumbre, *antes* de que los datos estén disponibles, cada observación se considera como una variable aleatoria y la muestra se denota por $X_1, X_2, \ldots, X_n$ (letras mayúsculas para variables aleatorias).

Esta variación en los valores observados implica a su vez que el valor de cualquier función de las observaciones muestrales, tal como la media muestral, la desviación estándar muestral o la dispersión de los cuartos muestrales, también varía de una muestra a otra. Es decir, antes de obtener $x_1, \ldots, x_n$, existe incertidumbre en cuanto al valor de $\bar{x}$, el valor de s, y así sucesivamente.

Ejemplo 5.19 Suponga que la resistencia del material de un espécimen seleccionado al azar de un tipo particular tiene una distribución Weibull con valores de parámetro $\alpha = 2$ (forma) y $\beta = 5$ (escala). La curva de densidad correspondiente se muestra en la figura 5.6. Las fórmulas de la sección 4.5 dan

$$\mu = E(x) = 4.4311 \qquad \tilde{\mu} = 4.1628 \qquad \sigma^2 = V(X) = 5.365 \qquad \sigma = 2.316$$

La media excede a la mediana debido a la asimetría positiva de la distribución.

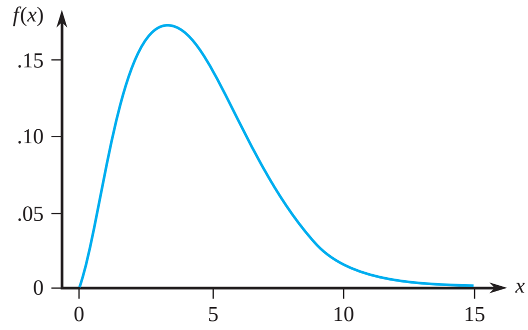


Figura 5.6 Curva de densidad Weibull del ejemplo 5.19

Se utilizó software estadístico para generar seis muestras diferentes, cada una con $n = 10$, de esta distribución (resistencias de material de seis diferentes grupos de diez especímenes cada uno). Los resultados aparecen en la tabla 5.1, seguidos por los valores de la media, la mediana y la desviación estándar de cada muestra. Obsérvese en primer lugar que las diez observaciones en cualquier muestra particular son diferentes de aquellas en cualquier otra muestra. En segundo lugar, los seis valores de la media muestral son diferentes entre sí, como lo son los seis valores de la mediana muestral y los seis valores de la desviación estándar de la muestra. Lo mismo es cierto para las medias recortadas 10%, la dispersiones de los cuartos de las muestras, y así sucesivamente.

Tabla 5.1 Muestras de la distribución Weibull del ejemplo 5.19

Muestra	1	2	3	4	5	6
1	6.1171	5.07611	3.46710	1.55601	3.12372	8.93795
2	4.1600	6.79279	2.71938	4.56941	6.09685	3.92487
3	3.1950	4.43259	5.88129	4.79870	3.41181	8.76202
4	0.6694	8.55752	5.14915	2.49759	1.65409	7.05569
5	1.8552	6.82487	4.99635	2.33267	2.29512	2.30932
6	5.2316	7.39958	5.86887	4.01295	2.12583	5.94195
7	2.7609	2.14755	6.05918	9.08845	3.20938	6.74166
8	10.2185	8.50628	1.80119	3.25728	3.23209	1.75468
9	5.2438	5.49510	4.21994	3.70132	6.84426	4.91827
10	4.5590	4.04525	2.12934	5.50134	4.20694	7.26081
$\bar{x}$	4.401	5.928	4.229	4.132	3.620	5.761
$\tilde{x}$	4.360	6.144	4.608	3.857	3.221	6.342
s	2.642	2.062	1.611	2.124	1.678	2.496

Además, el valor de la media de cualquier muestra puede ser considerado como *estimación puntual* ("puntual" porque es un solo número, correspondiente a un solo punto sobre la línea de numeración) de la media de la población μ, cuyo valor se sabe que es 4.4311. Ninguna de las estimaciones de estas seis muestras es idéntica a la que se está estimando. Las estimaciones de la segunda y sexta muestras son demasiado grandes, en tanto que la quinta da una subestimación sustancial. Asimismo, la desviación estándar muestral da una estimación puntual de la desviación estándar de la población. Las seis estimaciones resultantes están equivocadas por lo menos en una pequeña cantidad.

En resumen, los valores de las observaciones de las muestras individuales varían de una muestra a otra por lo que, en general, el valor de cualquier cantidad calculada a partir de datos de la muestra y el valor de una característica de la muestra utilizada como una estimación de la característica de la población correspondiente casi nunca coinciden con lo que se estimó. ■

DEFINICIÓN

Un **estadístico** es cualquier cantidad cuyo valor puede ser calculado a partir de datos muestrales. Antes de obtener los datos, existe incertidumbre sobre qué valor de cualquier estadístico particular resultará. Por consiguiente, un estadístico es una variable aleatoria y será denotado por una letra mayúscula; para representar el valor calculado u observado del estadístico se utiliza una letra minúscula.

Por lo tanto la media muestral, considerada como estadístico (antes de seleccionar una muestra o realizar un experimento), está denotada por $\overline{X}$; el valor calculado de este estadístico es $\overline{x}$. Del mismo modo, S representa la desviación estándar muestral considerada como estadístico y su valor calculado es s. Si se seleccionan muestras de dos tipos diferentes de ladrillos y las resistencias a la compresión individuales se denotan por $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$, respectivamente, entonces el estadístico $\overline{X} - \overline{Y}$, la diferencia entre las dos resistencias a la compresión muestrales medias, a menudo es de gran interés.

Cualquier estadístico, por el hecho de ser una variable aleatoria, tiene una distribución de probabilidad. En particular, la media muestral $\overline{X}$ tiene una distribución de probabilidad. Supóngase, por ejemplo, que $n = 2$ componentes se seleccionan al azar y que el número de descomposturas mientras se encuentran dentro de garantía se determina para cada uno. Los valores posibles del número medio muestral de descomposturas $\overline{X}$ son 0 (si $X_1 = X_2 = 0$), .5 (si $X_1 = 0$ y $X_2 = 1$ o $X_1 = 1$ y $X_2 = 0$), 1, 1.5, . . . La distribución de probabilidad de $\overline{X}$ especifica $P(\overline{X} = 0)$, $P(\overline{X} = .5)$ y así sucesivamente, a partir de las cuales otras probabilidades tales como $P(1 \le \overline{X} \le 3)$ y $P(\overline{X} \ge 2.5)$ pueden ser calculadas. Asimismo, si para una muestra de tamaño $n = 2$, los únicos valores posibles de la varianza muestral son 0, 12.5 y 50 (el cual es el caso si X_1 y X_2 pueden tomar sólo los valores 40, 45 o 50), entonces la distribución de probabilidad de S^2 da $P(S^2 = 0)$, $P(S^2 = 12.5)$ y $P(S^2 = 50)$. La distribución de probabilidad de un estadístico en ocasiones se conoce como **distribución de muestreo** para enfatizar que describe cómo varía el valor del estadístico a través de todas las muestras que pudieran ser seleccionadas.

Muestras aleatorias

La distribución de probabilidad de cualquier estadístico particular depende no sólo de la distribución de la población (normal, uniforme, etc.) y el tamaño de muestra n sino también del método de muestreo. Considérese seleccionar una muestra de tamaño $n = 2$ de una población compuesta de sólo los tres valores 1, 5 y 10, y supóngase que el estadístico de interés es la varianza muestral. Si el muestreo se realiza "con reemplazo", entonces $S^2 = 0$ resultará si $X_1 = X_2$. Sin embargo, S^2 no puede ser igual a 0 si el muestreo se realiza "sin reemplazo". Por tanto $P(S^2 = 0) = 0$ con un método de muestreo y esta probabilidad es positiva con el otro método. La siguiente definición describe un método de muestreo encontrado a menudo (por lo menos aproximadamente) en la práctica.

DEFINICIÓN

Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ forman una **muestra aleatoria (**simple) de tamaño n si

1. Las X_i son variables aleatorias independientes.

2. Cada X_i tiene la misma distribución de probabilidad.

Las condiciones 1 y 2 pueden ser parafraseadas diciendo que las X_i son *independientes e idénticamente distribuidas* (iid). Si el muestreo se realiza con reemplazo o de una población infinita (conceptual), las condiciones 1 y 2 se satisfacen con exactitud. Estas condiciones serán aproximadamente satisfechas si el muestreo se realiza sin reemplazo, aunque el tamaño n de la muestra es mucho más pequeño que el tamaño de la población N. En la práctica, si $n/N \le .05$ (cuando mucho 5% de la población se muestrea), se puede proceder como si las X_i formaran una muestra aleatoria. La virtud de este método de muestreo es que la distribución de probabilidad de cualquier estadístico es más fácil de obtener que con cualquier otro método de muestreo.

Existen dos métodos generales para obtener información sobre una distribución de muestreo de un estadístico. Uno implica cálculos basados en reglas de probabilidad y el otro implica realizar un experimento de simulación.

Deducción de una distribución de muestreo

Se pueden utilizar reglas de probabilidad para obtener la distribución de un estadístico siempre que sea una función "bastante simple" de las X_i y existen relativamente pocos valores X diferentes en la población o bien la distribución de la población tiene una forma "accesible". Los dos ejemplos siguientes ilustran tales situaciones.

Ejemplo 5.20 Una cierta marca de reproductor MP3 viene en tres configuraciones: un modelo con 2 GB de memoria, que cuesta $80, un modelo de 4 GB a un precio de $100 y una versión de 8 GB con un precio de $120. Si el 20% de todos los compradores elige el modelo de 2 GB, el 30% elige el modelo de 4 GB, y el 50% elige el modelo de 8 GB, entonces la distribución de probabilidad del costo X de una sola compra de un reproductor de MP3 seleccionado al azar está dada por

x	80	100	120
$p(x)$	.2	.3	.5

con $\mu = 106$, $\sigma^2 = 244$ (5.2)

Supongamos que en un día particular sólo se venden dos reproductores de MP3. Sean X_1 = los ingresos procedentes de la primera venta y X_2 = los ingresos de la segunda. Supongamos que X_1 y X_2 son independientes, cada uno con la distribución de probabilidad que se muestra en (5.2) [de manera que X_1 y X_2 constituyen una muestra aleatoria de la distribución (5.2)]. En la Tabla 5.2 se enumeran posibles pares (x_1, x_2), la probabilidad de cada uno [calculada utilizando (5.2) y la suposición de independencia] y los valores resultantes de $\bar{x}$ y s^2. [Tenga en cuenta que cuando $n = 2$, $s^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$.] Ahora para obtener la distribución de probabilidad $\bar{X}$, la muestra de los ingresos promedio por venta, tenemos que considerar cada posible valor de $\bar{x}$ y calcular su probabilidad. Por ejemplo, $\bar{x} = 100$ aparece tres veces en la tabla con probabilidades de .10, .09 y .10, por lo que

$$p_{\bar{X}}(100) = P(\bar{X} = 100) = .10 + .09 + .10 = .29$$

Asimismo,

$$\begin{aligned} pS^2(800) = P(S^2 = 800) &= P(X_1 = 80, X_2 = 120 \text{ o } X_1 = 120, X_2 = 80) \\ &= .10 + .10 = .20 \end{aligned}$$

Tabla 5.2 Resultados, probabilidades y valores de $\bar{x}$ y s^2 en el ejemplo 5.20

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	.04	80	0
80	100	.06	90	200
80	120	.10	100	800
100	80	.06	90	200
100	100	.09	100	0
100	120	.15	110	200
120	80	.10	100	800
120	100	.15	110	200
120	120	.25	120	0

Las distribuciones de muestreo completas de $\overline{X}$ y S^2 aparecen en (5.3) y (5.4).

$\bar{x}$	80	90	100	110	120
$p_{\overline{X}}(\bar{x})$	.04	.12	.29	.30	.25

(5.3)

s^2	0	200	800
$pS^2(s^2)$	.38	.42	.20

(5.4)

La figura 5.7 ilustra un histograma de probabilidad tanto de la distribución original (5.2) como de la distribución $\overline{X}$ (5.3). La figura sugiere primero que la media (valor esperado) de la distribución $\overline{X}$ es igual a la media 106 de la distribución original, puesto que ambos histogramas parecen estar centrados en el mismo lugar.
De acuerdo con (5.3),

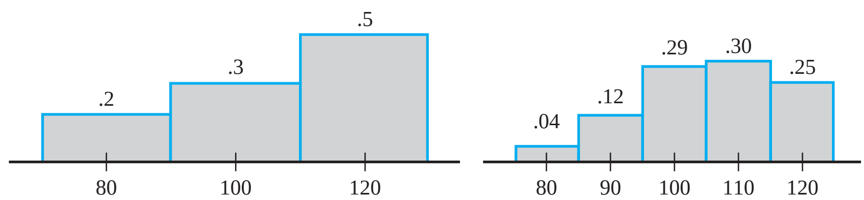


Figura 5.7 Histogramas de probabilidad de la distribución subyacente y distribución $\overline{X}$, en el ejemplo 5.20

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x} p_{\overline{X}}(\bar{x}) = (80)(.04) + \cdots + (120)(.25) = 106 = \mu$$

En segundo lugar, parece que la distribución $\overline{X}$ tiene una dispersión más pequeña (variabilidad) que la distribución original, puesto que la masa de probabilidad se movió hacia la media. De nuevo de acuerdo con (5.3),

$$\begin{aligned}\sigma_{\overline{X}}^2 &= V(\overline{X}) = \sum \bar{x}^2 \cdot p_{\overline{X}}(\bar{x}) - \mu_{\overline{X}}^2 \\ &= (80^2)(.04) + \cdots + (120^2)(.25) - (106)^2 \\ &= 122 = \frac{244}{2} = \frac{\sigma^2}{2}\end{aligned}$$

La varianza de $\overline{X}$ es precisamente la mitad de la varianza original (porque $n = 2$).

Utilizando (5.4) el valor medio de S^2 es

$$\begin{aligned}\mu S^2 &= E(S^2) = \sum S^2 \cdot p_{S^2}(s^2) \\ &= (0)(.38) + (200)(.42) + (800)(.20) = 244 = \sigma^2\end{aligned}$$

Es decir, la distribución de muestreo $\overline{X}$ tiene su centro en la media de la población μ y la distribución de muestreo S^2 está centrada en la varianza de la población σ^2.

Si se hubieran realizado cuatro compras en el día de interés, el ingreso promedio muestral $\overline{X}$ estaría basado en una muestra aleatoria de cuatro X_i, cada una con la distribución (5.2). Más cálculo a la larga da la función de masa de probabilidad de $\overline{X}$ para $n = 4$ como

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\overline{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625

De acuerdo con esto, $\mu_{\overline{X}} = 106 = \mu$ y $\sigma^2_{\overline{X}} = 61 = \sigma^2/4$. La figura 5.8 es un histograma de probabilidad de esta función de masa de probabilidad.

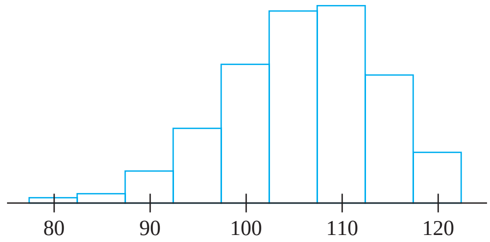


Figura 5.8 Histograma de probabilidad de $\overline{X}$ basado en $n = 4$ en el ejemplo 5.20

El ejemplo 5.20 sugiere primero que todo que los cálculos de $p_{\overline{X}}(\overline{x})$ y $p_{S^2}(s^2)$ pueden ser tediosos. Si la distribución original (5.2) hubiera permitido más de tres valores posibles, entonces incluso con $n = 2$ los cálculos hubieran sido más complicados. El ejemplo también sugiere, sin embargo, que existen algunas relaciones generales entre $E(\overline{X})$, $V(\overline{X})$, $E(S^2)$ y la media μ y la varianza σ^2 de la distribución original. Éstas se formulan en la siguiente sección. Ahora considérese un ejemplo en el cual la muestra aleatoria se saca de una distribución continua.

Ejemplo 5.21 El tiempo de servicio para un tipo de transacción bancaria es una variable aleatoria con distribución exponencial y parámetro λ. Suponga que X_1 y X_2 son tiempos de servicio para dos clientes diferentes, supuestos independientes entre sí. Considere el tiempo de servicio total $T_o = X_1 + X_2$ para los dos clientes, también un estadístico. La función de distribución acumulativa de T_o con $t \geq 0$, es

$$F_{T_0}(t) = P(X_1 + X_2 \leq t) = \iint_{\{(x_1, x_2):\, x_1 + x_2 \leq t\}} f(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_0^t \int_0^{t - x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1 = \int_0^t [\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}]\, dx_1$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$

La región de integración se ilustra en la figura 5.9.

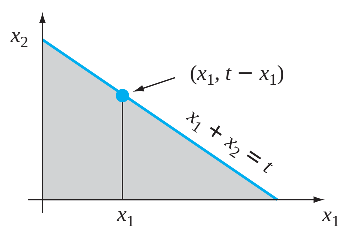


Figura 5.9 Región de integración para obtener la función de distribución acumulativa de T_o en el ejemplo 5.21

La función de densidad de probabilidad de T_o se obtiene diferenciando $F_{T_o}(t)$:

$$f_{T_o}(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{5.5}$$

Ésta es una función de densidad de probabilidad gamma ($\alpha = 2$ y $\beta = 1/\lambda$). La función de densidad de probabilidad de $\overline{X} = T_o/2$ se obtiene a partir de la relación $\{\overline{X} \le \overline{x}\}$ si y sólo si $\{T_o \le 2\overline{x}\}$ como

$$f_{\overline{X}}(\overline{x}) = \begin{cases} 4\lambda^2 \overline{x} e^{-2\lambda\overline{x}} & \overline{x} \ge 0 \\ 0 & \overline{x} < 0 \end{cases} \tag{5.6}$$

La media y la varianza de la distribución exponencial subyacente son $\mu = 1/\lambda$ y $\sigma^2 = 1/\lambda^2$. Con las expresiones (5.5) y (5.6) se puede verificar que $E(\overline{X}) = 1/\lambda$, $V(\overline{X}) = 1/(2\lambda^2)$, $E(T_o) = 2/\lambda$ y $V(T_o) = 2/\lambda^2$. Estos resultados sugieren de nuevo algunas relaciones generales entre medias y varianzas de $\overline{X}$, T_o y la distribución subyacente.

Experimentos de simulación

El segundo método de obtener información sobre distribución de muestreo de un estadístico es realizar un experimento de simulación. Este método casi siempre se utiliza cuando una derivación vía reglas de probabilidad es demasiado difícil o complicada de realizar. Tal experimento virtualmente se realiza siempre con la ayuda de una computadora. Las siguientes características de un experimento deben ser especificadas:

1. El estadístico de interés ($\overline{X}$, S, una media recortada particular, etc.)
2. La distribución de la población (normal con $\mu = 100$ y $\sigma = 15$, uniforme con límite inferior $A = 5$ y superior $B = 10$, etc.)
3. El tamaño de muestra n (p. ej., $n = 10$ o $n = 50$)
4. El número de réplicas k (número de muestras que serán obtenidas)

Luego se utiliza una computadora para obtener k diferentes muestras aleatorias, cada una de tamaño n, de la distribución de población designada. Para cada una de las muestras, calcule el valor del estadístico y construya un histograma de los k valores. Este histograma da la distribución de muestreo *aproximada* del estadístico. Mientras más grande es el valor de k, mejor tenderá a ser la aproximación (la distribución de muestreo real emerge a medida que $k \to \infty$). En la práctica, $k = 500$ o 1000 casi siempre es suficiente si el estadístico es "bastante simple".

Ejemplo 5.22 La distribución de la población del primer estudio de simulación es normal con $\mu = 8.25$ y $\sigma = .75$, como se ilustra en la figura 5.10. [El artículo "Platelet Size in Myocardial Infarction" (*British Med. J.*, 1983: 449–451) sugiere esta distribución de volumen de plaquetas en individuos sin historial clínico de problemas cardiacos serios.]

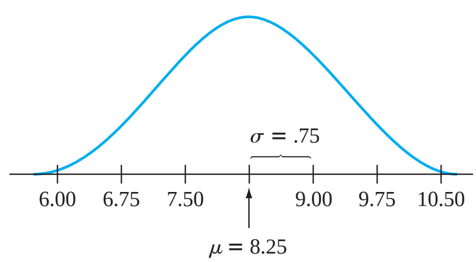


Figura 5.10 Distribución normal, con $\mu = 8.25$ y $\sigma = .75$

En realidad se realizaron cuatro experimentos diferentes, con 500 réplicas por cada uno. En el primero, se generaron 500 muestras de $n = 5$ observaciones cada una con Minitab y los tamaños de las otras tres muestras fueron $n = 10$, $n = 20$ y $n = 30$, respectivamente. La media muestral se calculó para cada muestra, y los histogramas resultantes de valores $\bar{x}$ aparecen en la figura 5.11.

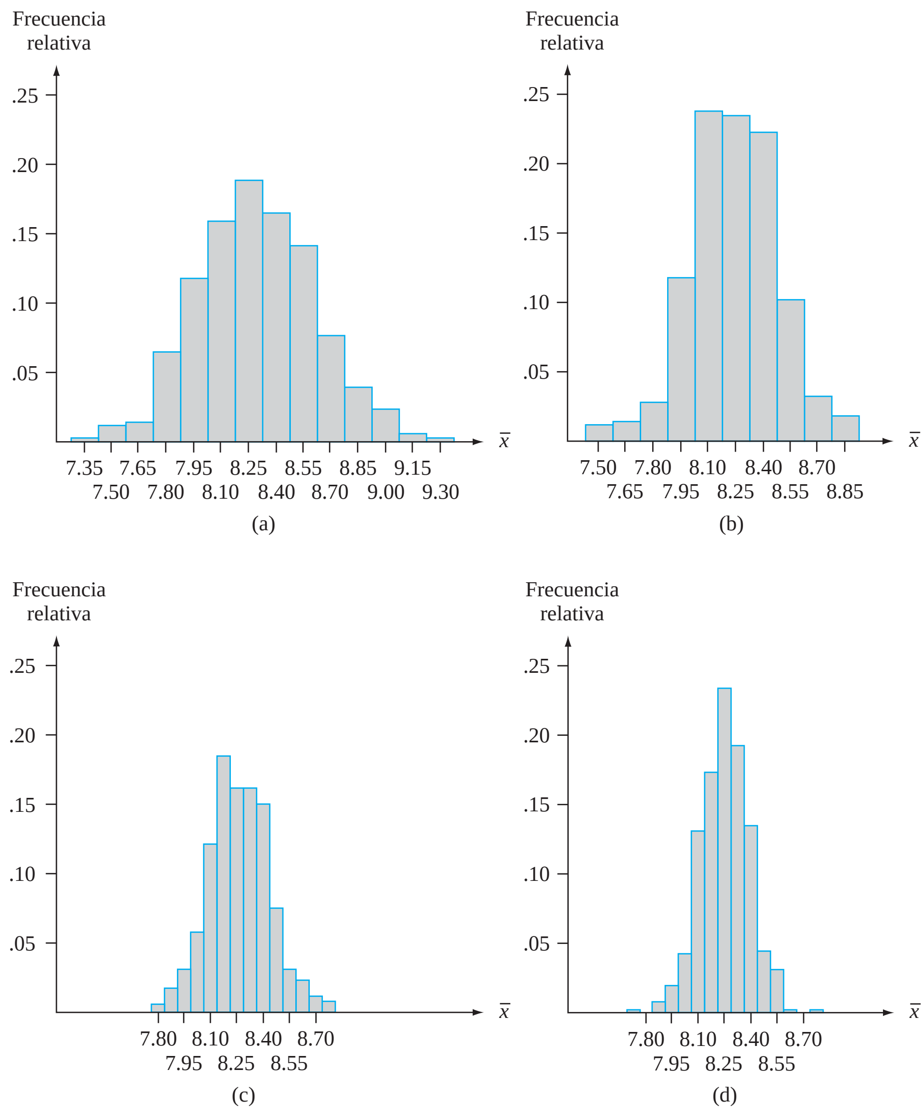


Figura 5.11 Histogramas muestrales de $\bar{x}$ basados en 500 muestras, cada una compuesto de n observaciones: (a) $n = 5$; (b) $n = 10$; (c) $n = 20$; (d) $n = 30$

Lo primero que se nota en relación con los histogramas es su forma. Con una razonable aproximación, cada uno de los cuatro se ve como una curva normal. El parecido sería aún más impactante si cada histograma se hubiera basado en mucho más que 500 valores $\bar{x}$. En segundo lugar, cada histograma está centrado aproximadamente en 8.25, la media de la población muestreada. Si los histogramas se hubieran basado en una secuencia interminable de valores $\bar{x}$, sus centros habrían sido exactamente la media de la población, 8.25.

El aspecto final del histograma es su dispersión uno con respecto al otro. Mientras más grande es el valor de n, más concentrada está la distribución muestral sobre el valor medio. Por eso los histogramas con $n = 20$ y $n = 30$ están basados en intervalos de clase más angostos que aquellos para los dos tamaños de muestra más pequeños. Con los tamaños de muestra más grandes, la mayoría de los valores $\bar{x}$ están bastante cerca de 8.25. Éste es el efecto de promediar. Cuando n es pequeño, un solo valor x inusual puede dar por resultado un valor $\bar{x}$ alejado del centro. Con un tamaño de muestra grande, cualesquiera valores x inusuales, cuando se promedian con los demás valores muestrales, seguirán tendiendo a producir un valor $\bar{x}$ próximo a μ. Si se combinan estas ideas se obtiene un resultado muy apegado a su intuición: **$\bar{X}$ basado en n grande tiende a acercarse más a μ que $\bar{X}$ basado en n pequeño.**

Ejemplo 5.23 Considere un experimento de simulación en el cual la distribución de la población es bastante asimétrica. La figura 5.12 muestra la curva de densidad de las vidas útiles de un tipo de control electrónico [ésta es en realidad una distribución lognormal con $E(\ln(X)) = 3$ y $V(\ln(X)) = .16$]. De nueva cuenta el estadístico de interés es la media muestral $\bar{X}$. El experimento utilizó 500 réplicas y consideró los mismos cuatro tamaños de muestra que en el ejemplo 5.22. Los histogramas resultantes junto con una curva de probabilidad normal generada por Minitab con los 500 valores $\bar{x}$ basados en $n = 30$ se muestran en la figura 5.13.

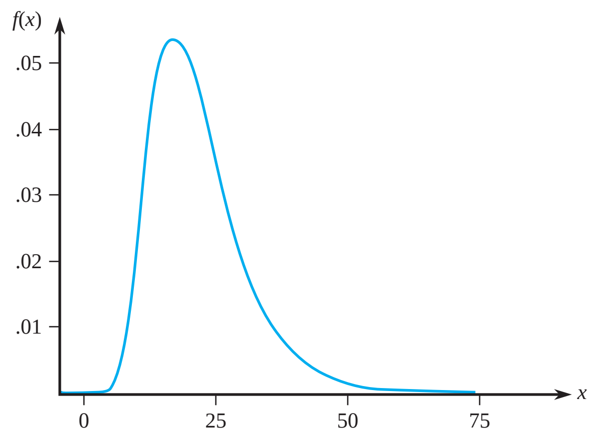


Figura 5.12 Curva de densidad del experimento de simulación del ejemplo 5.23 [$E(X) = 21.7584$, $V(X) = 82.1449$]

A diferencia del caso normal, estos histogramas difieren en cuanto a forma. En particular, se vuelven progresivamente menos asimétricos a medida que el tamaño de muestra n se incrementa. El promedio de los 500 valores $\bar{x}$ con los cuatro tamaños de muestra diferentes se aproximan bastante al valor medio de la distribución de la población. Si cada histograma se hubiera basado en una secuencia interminable de valores $\bar{x}$ en lugar de en sólo 500, los cuatro habrían tenido su centro en exactamente 21.7584. Por tanto, los valores diferentes de n cambian la forma mas no el centro de la distribución de muestreo de $\bar{X}$. La comparación de los cuatro histogramas en la figura 5.13 también muestra que conforme n se incrementa, la dispersión de los histogramas decrece. El incremento de n produce un mayor grado de concentración en torno al valor medio de la población y hace que el histograma se vea más como una curva normal. El histograma de la figura 5.13(d) y la curva de probabilidad normal en la figura 5.13(e) proporcionan una evidencia convincente de que un tamaño de muestra de $n = 30$ es suficiente para superar la asimetría de la distribución de la población y para producir una distribución de muestreo $\bar{X}$ aproximadamente normal.

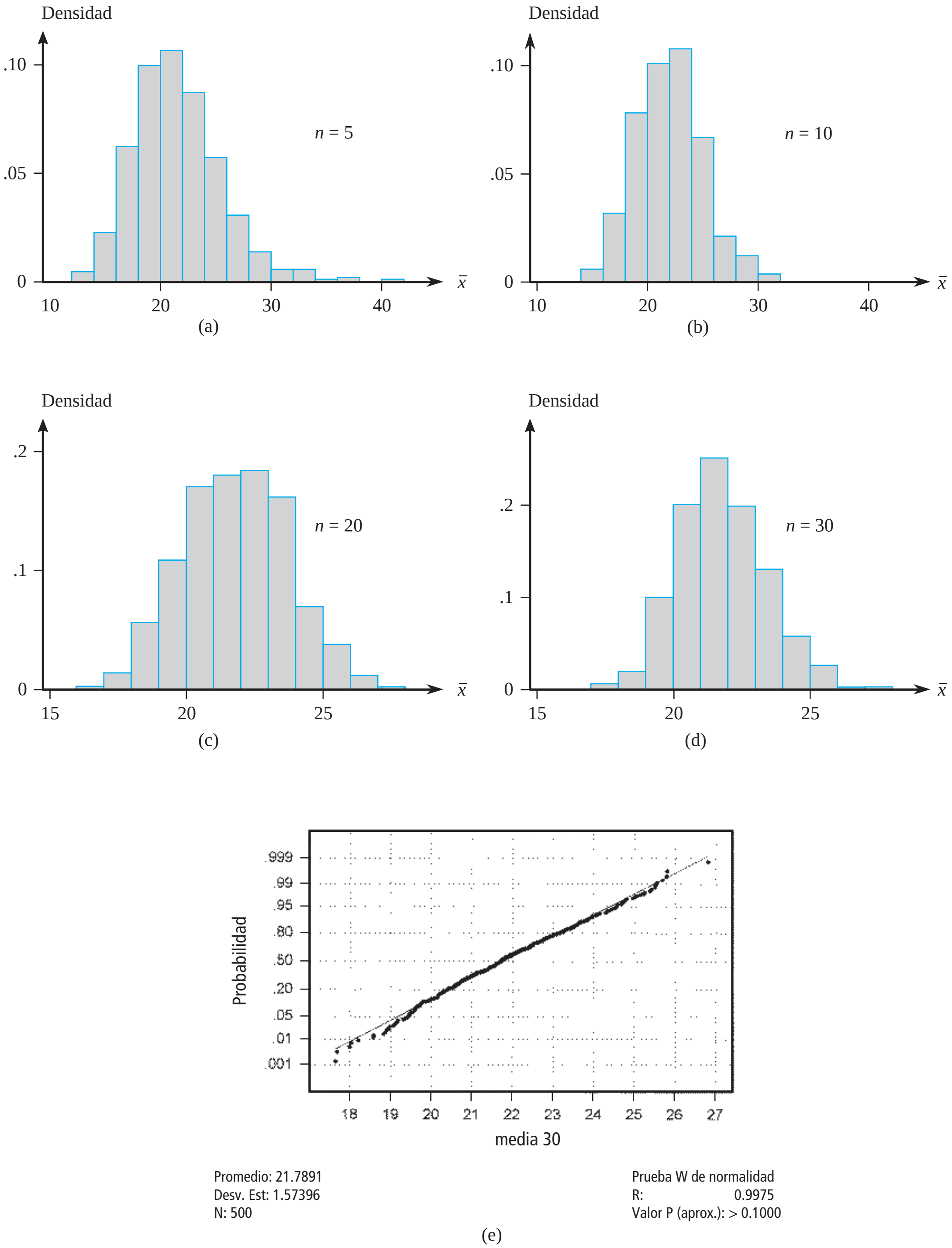


Figura 5.13 Resultados del experimento de simulación del ejemplo 5.23: (a) histograma de $\bar{x}$ con $n = 5$; (b) histograma de $\bar{x}$ con $n = 10$; (c) histograma de $\bar{x}$ con $n = 20$; (d) histograma de $\bar{x}$ con $n = 30$; (e) curva de probabilidad normal con $n = 30$ (generados por Minitab)

EJERCICIOS Sección 5.3 (37–45)

37. Una marca particular de jabón para lavadora de platos se vende en tres tamaños: 25 oz, 40 oz y 65 oz. Veinte por ciento de todos los compradores seleccionan la caja de 25 oz, 50% seleccionan una caja de 40 oz y el 30% restante seleccionan la caja de 65 oz. Sean X_1 y X_2 los tamaños de paquete seleccionados por dos compradores independientemente seleccionados.

a. Determine la distribución de muestreo de $\overline{X}$, calcule $E(\overline{X})$ y compare con μ.

b. Determine la distribución de muestreo de la varianza muestral S^2, calcule $E(S^2)$ y compare con σ^2.

38. Hay dos semáforos en mi camino de ida y vuelta al trabajo. Sea X_1 el número de semáforos en los cuales me tengo que detener a la ida al trabajo y sea X_2 el número de semáforos en los cuales me tengo que detener de regreso a casa. Suponga que estas dos variables son independientes cada una con una fmp dada en la tabla adjunta (de modo que X_1, X_2 es una muestra aleatoria de tamaño $n = 2$).

x_1	0	1	2
$p(x_1)$	.2	.5	.3

$\mu = 1.1,\ \sigma^2 = .49$

a. Sea $T_o = X_1 + X_2$ y determine la función de masa de probabilidad.

b. Calcule μ_{T_o}. ¿Cómo se relaciona con μ, la media de la población?

c. Calcule $\sigma^2_{T_o}$. ¿Cómo se relaciona con σ^2, la varianza de la población?

d. Sean X_3 y X_4 el número de luces en las que se requiere una parada durante la conducción hacia el trabajo y de regreso por segundo día consecutivo, suponiendo que es independiente de la primera jornada. Con T_0 = la suma de todos los cuatro X_i, ¿cuáles son ahora los valores de $E(T_0)$ y $V(T_0)$?

e. Refiriéndose de nuevo a (d), ¿cuáles son los valores de $P(T_o = 8)$ y $P(T_o \geq 7)$? [*Sugerencia*: no se te ocurra incluir todos los resultados posibles!]

39. Se sabe que 80% de todos los discos de almacenamiento extraíbles funcionan satisfactoriamente durante el periodo de garantía (son “éxitos”). Suponga que se seleccionan al azar $n = 10$ unidades de disco. Sea X = el número de éxitos en la muestra. El estadístico X/n es la proporción de la muestra (fracción) de éxitos. Obtenga la distribución muestral de este estadístico. [*Sugerencia*: un posible valor de X/n es .3, correspondiente a $X = 3$. ¿Cuál es la probabilidad de este valor (qué clase de variable aleatoria es X)?]

40. Una caja contiene diez sobres sellados numerados 1, . . . , 10. Los primeros cinco no contienen dinero, cada uno de los siguientes tres contiene \$5 y hay un billete de \$10 en cada uno de los últimos dos. Se selecciona un tamaño de muestra de 3 *con* reemplazo (así que se tiene una muestra aleatoria) y se obtiene la cantidad más grande en cualquiera de los sobres seleccionados. Si X_1, X_2 y X_3 denotan las cantidades en los sobres seleccionados, el estadístico de interés es M = el máximo de X_1, X_2 y X_3.

a. Obtenga la distribución de probabilidad de este estadístico.

b. Describa cómo realizaría un experimento de simulación para comparar las distribuciones de M con varios tamaños de muestra. ¿Cómo piensa que cambiaría la distribución a medida que n se incrementa?

41. Sea X el número de paquetes enviados por un cliente seleccionado al azar vía una compañía de paquetería y mensajería. Suponga que la distribución de X es como sigue:

x	1	2	3	4
$p(x)$	.4	.3	.2	.1

a. Considere una muestra aleatoria de tamaño $n = 2$ (dos clientes) y sea $\overline{X}$ el número medio muestral de paquetes enviados. Obtenga la distribución de probabilidad de $\overline{X}$.

b. Remítase al inciso (a) y calcule $P(\overline{X} \leq 2.5)$.

c. De nuevo considere una muestra aleatoria de tamaño $n = 2$, pero ahora enfóquese en el estadístico R = el rango muestral (diferencia entre los valores más grande y más pequeño en la muestra). Obtenga la distribución de R. [*Sugerencia*: calcule el valor de R para cada resultado y use las probabilidades del inciso (a).]

d. Si se selecciona una muestra aleatoria de tamaño $n = 4$, ¿cuál es $P(\overline{X} \leq 1.5)$? [*Sugerencia*: no tiene que dar todos los resultados posibles, sólo aquellos para los cuales $\bar{x} \leq 1.5$.]

42. Una compañía mantiene tres oficinas en una región, cada una manejada por dos empleados. Información concerniente a salarios anuales (miles de dólares) es la siguiente:

Oficina	1	1	2	2	3	3
Empleado	1	2	3	4	5	6
Salario	29.7	33.6	30.2	33.6	25.8	29.7

a. Suponga que dos de estos empleados se seleccionan al azar de entre los seis (sin reemplazo). Determine la distribución muestral del salario medio muestral $\overline{X}$.

b. Suponga que se selecciona al azar una de las tres oficinas. Sean X_1 y X_2 los salarios de los dos empleados. Determine la distribución muestral de $\overline{X}$.

c. ¿Cómo se compara $E(\overline{X})$ de los incisos (a) y (b) con el salario medio de la población μ?

43. Suponga que la cantidad de líquido despachada por una máquina está uniformemente distribuida con límite inferior $A = 8$ oz y límite superior $B = 10$ oz. Describa cómo realizaría experimentos de simulación para comparar la distribución muestral de la dispersión de los cuartos (muestral) con tamaños de muestra $n = 5$, 10, 20 y 30.

44. Realice un experimento de simulación con un paquete de computadora estadístico u otro programa para estudiar la distribución muestral de $\overline{X}$ cuando la distribución de la población es de Weibull con $\alpha = 2$ y $\beta = 5$, como en el ejemplo 5.19. Considere los cuatro tamaños de muestra $n = 5$, 10, 20 y 30, y en cada caso utilice 1000 réplicas. ¿Con cuál de estos tamaños de muestra la distribución muestral $\overline{X}$ parece ser aproximadamente normal?

45. Realice un experimento de simulación con un paquete de computadora estadístico u otro programa para estudiar la distribución muestral de $\overline{X}$ cuando la distribución de la población es lognormal con $E(\ln(X)) = 3$ y $V(\ln(X)) = 1$. Considere los cuatro tamaños de muestra $n = 10$, 20, 30 y 50 y en cada caso utilice 1000 réplicas. ¿Con cuál de estos tamaños de muestra la distribución muestral $\overline{X}$ parece ser aproximadamente normal?

5.4 Distribución de la media muestral

La importancia de la media muestral $\overline{X}$ proviene de su uso al sacar conclusiones sobre la media de la población μ. Algunos de los procedimientos inferenciales más frecuentemente utilizados están basados en propiedades de la distribución muestral de $\overline{X}$. Un examen previo de estas propiedades apareció en los cálculos y experimentos de simulación de la sección previa, donde se observaron las relaciones entre $E(\overline{X})$ y μ y también entre $V(\overline{X})$, σ^2 y n.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con valor medio μ y desviación estándar σ. Entonces

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$

2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ y $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

Además, con $T_o = X_1 + \cdots + X_n$ (el total de la muestra), $E(T_o) = n\mu$, $V(T_o) = n\sigma^2$ y $\sigma_{T_o} = \sqrt{n}\sigma$.

Las demostraciones de estos resultados se difieren a la siguiente sección. De acuerdo con el resultado 1, la distribución (es decir, probabilidad) muestral de $\overline{X}$ está centrada precisamente en la media de la población de la cual se seleccionó la muestra. El resultado 2 muestra que la distribución $\overline{X}$ se concentra más en torno a μ a medida que se incrementa el tamaño n de la muestra. En un marcado contraste, la distribución de T_o se dispersa más a medida que n se incrementa. Al promediar la probabilidad se mueve hacia la parte media, en tanto que al totalizar la probabilidad se dispersa sobre un rango más y más amplio de valores. La desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ es a menudo llamada *error estándar de la media*; éste describe la magnitud de una desviación típica o representativa de la media muestral respecto de la media poblacional.

Ejemplo 5.24 En una prueba de fatiga por tensión con un espécimen de titanio, el número esperado de ciclos hasta la primera emisión acústica (utilizada para indicar la iniciación del agrietamiento) es $\mu = 28{,}000$, y la desviación estándar del número de ciclos es $\sigma = 5000$. Sea $X_1, X_2, \ldots, X_{25}$ una muestra aleatoria de tamaño 25, donde cada X_i es el número de ciclos en un espécimen diferente seleccionado al azar. Entonces el valor esperado de la media muestral del número de ciclos hasta la primera emisión es $E(\overline{X}) = \mu = 28{,}000$ y el número total esperado de ciclos para los 25 especímenes es $E(T_0) = n\mu = 25(28{,}000) = 700{,}000$. La desviación estándar de $\overline{X}$ (error estándar de la media) y de T_o son

$$\sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{5000}{\sqrt{25}} = 1000$$

$$\sigma_{T_o} = \sqrt{n}\sigma = \sqrt{25}(5000) = 25{,}000$$

Si el tamaño de la muestra se incrementa a $n = 100$, $E(\overline{X})$ no cambió, pero $\sigma_{\overline{X}} = 500$, la mitad de su valor previo (el tamaño de muestra debe ser cuadruplicado para reducir a la mitad la desviación estándar de $\overline{X}$). ■

El caso de una distribución de población normal

El experimento de simulación del ejemplo 5.22 indicó que cuando la distribución de la población es normal, cada histograma de valores $\overline{x}$ se aproxima muy bien con una curva normal.

PROPOSICIÓN

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución *normal* con media μ y desviación estándar σ. Entonces con *cualquier n*, $\overline{X}$ está normalmente distribuida (con media μ y desviación estándar $\sigma/\sqrt{n}$), como T_o (con media $n\mu$ y desviación estándar $\sqrt{n}\sigma$).*

Se sabe todo lo que se tiene que saber sobre las distribuciones $\overline{X}$ y T_o cuando la distribución de la población es normal. En particular, probabilidades tales como $P(a \le \overline{X} \le b)$ y $P(c \le T_o \le d)$ se obtienen simplemente estandarizando. La figura 5.14 ilustra la proposición.

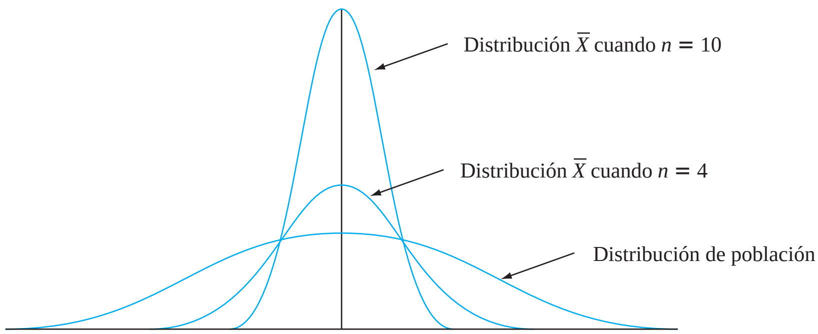


Figura 5.14 Distribución de población normal y distribuciones muestrales $\overline{X}$

Ejemplo 5.25 El tiempo que requiere una rata de cierta subespecie seleccionada al azar para encontrar su camino a través de un laberinto es una variable aleatoria normalmente distribuida con $\mu = 1.5$ min y $\sigma = .35$ min. Suponga que se seleccionan cinco ratas. Sean $X_1, \ldots, X_5$ sus tiempos en el laberinto. Suponiendo que las X_i son una muestra aleatoria de esta distribución normal, ¿cuál es la probabilidad de que el tiempo total $T_o = X_1 + \ldots + X_5$ de las cinco sea de entre 6 y 8 min? De acuerdo con la proposición, T_o tiene una distribución normal con $\mu_{T_o} = n\mu = 5(1.5) = 7.5$ y varianza $\sigma^2_{T_o} = n\sigma^2 = 5(.1225) = .6125$, por tanto $\sigma_{T_o} = .783$. Para estandarizar T_o, reste μ_{T_o} y divida entre σ_{T_o}:

$$P(6 \le T_o \le 8) = P\left(\frac{6 - 7.5}{.783} \le Z \le \frac{8 - 7.5}{.783}\right)$$
$$= P(-1.92 \le Z \le .64) = \Phi(.64) - \Phi(-1.92) = .7115$$

La determinación de la probabilidad de que el tiempo promedio muestral $\overline{X}$ (una variable normalmente distribuida) sea cuando mucho de 2.0 min requiere $\mu_{\overline{X}} = \mu = 1.5$ y $\sigma_{\overline{X}} = \sigma/\sqrt{n} = .35/\sqrt{5} = .1565$. Entonces

$$P(\overline{X} \le 2.0) = P\left(Z \le \frac{2.0 - 1.5}{.1565}\right) = P(Z \le 3.19) = \Phi(3.19) = .9993$$ ■

* Una prueba del resultado para T_o cuando $n = 2$ es posible si se utiliza el método del ejemplo 5.21, pero los detalles son complicados. El resultado general casi siempre se comprueba por medio de una herramienta teórica llamada *función generadora de momentos*. Se puede consultar una de las referencias del capítulo para más información.

Teorema del límite central

Cuando las X_i están normalmente distribuidas, también lo está $\overline{X}$ con cada tamaño de muestra n. Las deducciones del ejemplo 5.20 y el experimento de simulación del ejemplo 5.23, sugieren que incluso cuando la distribución de la población es altamente no normal, el cálculo de promedios produce una distribución más acampanada que la que está siendo muestreada. Una conjetura razonable es que si n es grande, una curva normal apropiada representará de forma más o menos aproximada la distribución real de $\overline{X}$. El planteamiento formal de este resultado es el más importante teorema de probabilidad.

TEOREMA

Teorema del límite central (TLC)

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces si n es suficientemente grande, $\overline{X}$ tiene aproximadamente una distribución normal con $\mu_{\overline{X}} = \mu$ y $\sigma^2_{\overline{X}} = \sigma^2/n$, y T_o también tiene aproximadamente una distribución normal con $\mu_{T_o} = n\mu$, $\sigma^2_{T_o} = n\sigma^2$. Mientras más grande es el valor de n, mejor es la aproximación.

La figura 5.15 ilustra el teorema del límite central. De acuerdo con el TLC, cuando n es grande y se desea calcular una probabilidad como $P(a \le \overline{X} \le b)$, lo único que se requiere es "aparentar" que $\overline{X}$ es normal, estandarizarla y utilizar la tabla normal. La respuesta resultante será aproximadamente correcta. Se podría obtener la respuesta correcta determinando primero la distribución de $\overline{X}$, así que el TLC proporciona un atajo verdaderamente impresionante. La comprobación del teorema implica muchas matemáticas avanzadas.

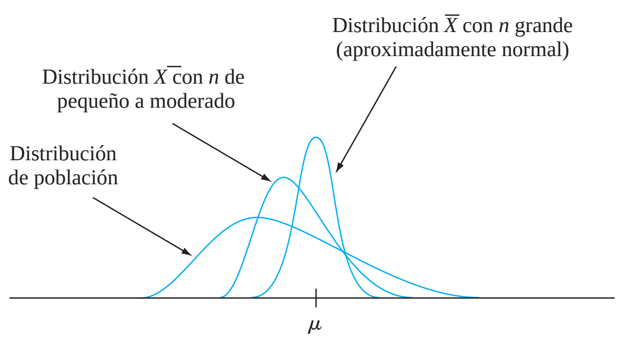


Figura 5.15 Teorema del límite central ilustrado

Ejemplo 5.26 La cantidad de una impureza particular en un lote de cierto producto químico es una variable aleatoria con valor medio de 4.0 g y desviación estándar de 1.5 g. Si se preparan 50 lotes en forma independiente, ¿cuál es la probabilidad (aproximada) de que la cantidad promedio muestral de la impureza $\overline{X}$ esté entre 3.5 a 3.8 g? De acuerdo con la regla empírica que se formulará en breve, $n = 50$ es suficientemente grande como para que el TLC sea aplicable. En ese caso $\overline{X}$ tiene aproximadamente una distribución normal con valor medio $\mu_{\overline{X}} = 4.0$ y $\sigma_{\overline{X}} = 1.5/\sqrt{50} = .2121$, por lo tanto

$$P(3.5 \le \overline{X} \le 3.8) \approx P\left(\frac{3.5 - 4.0}{.2121} \le Z \le \frac{3.8 - 4.0}{.2121}\right)$$
$$= \Phi(-.94) - \Phi(-2.36) = .1645$$ ■

Ejemplo 5.27 Una organización de protección al consumidor reporta cotidianamente el número de defectos mayores de cada automóvil nuevo que prueba. Suponga que el número de tales defectos en cierto modelo es una variable aleatoria con valor medio de 3.2 y desviación estándar de 2.4. Entre 100 carros seleccionados al azar de este modelo, ¿qué tan probable es que el número promedio muestral de defectos mayores exceda de 4? Sea X_i el número de defectos mayores del carro *i*ésimo en la muestra aleatoria. Obsérvese que X_i es una variable aleatoria discreta, pero el TLC es aplicable si la variable de interés es discreta o continua. Además, aunque el hecho de que la desviación estándar de esta variable no negativa es bastante grande con respecto al valor medio sugiere que su distribución es positivamente asimétrica, el gran tamaño de muestra implica que $\overline{X}$ sí tiene aproximadamente una distribución normal. Con $\mu_{\overline{X}} = 3.2$ y $\sigma_{\overline{X}} = .24$,

$$P(\overline{X} > 4) \approx P\left(Z > \frac{4 - 3.2}{.24}\right) = 1 - \Phi(3.33) = .0004$$ ■

El TLC da una idea de por qué muchas variables aleatorias tienen distribuciones de probabilidad que son aproximadamente normales. Por ejemplo, el error de medición en un experimento científico puede ser considerado como la suma de varias perturbaciones y errores subyacentes de pequeña magnitud.

Una dificultad práctica al aplicar el teorema del límite central es saber cuándo *n* es suficientemente grande. El problema es que la precisión de la aproximación con una *n* particular depende de la forma de la distribución subyacente original que está siendo muestreada. Si la distribución subyacente tiende a una curva de densidad normal, entonces la aproximación será buena incluso con *n* pequeña, mientras que si está lejos de ser normal, entonces se requerirá una *n* grande.

Regla empírica

Si $n > 30$, se puede utilizar el teorema del límite central.

Existen distribuciones de población para las cuales incluso una *n* de 40 o 50 no es suficiente, pero tales distribuciones rara vez se encuentran en la práctica. Por otra parte, la regla empírica a menudo es conservadora; para muchas distribuciones de población, una *n* mucho menor que 30 sería suficiente. Por ejemplo, en el caso de una distribución de población uniforme, el teorema del límite central da una buena aproximación con $n \geq 12$.

Ejemplo 5.28 Considere la distribución que se muestra en la figura 5.16 para la cantidad comprada (redondeada al dólar más cercano) por un cliente seleccionado al azar en una gasolinera en particular (una distribución similar para las compras en Gran Bretaña (en libras) apareció en el artículo "Data Mining for Fun and Profit", *Statistical Science*, 2000: 111−131; hubo grandes picos en los valores, 10, 15, 20, 25 y 30). La distribución es, obviamente, muy distinta de la normal.

Se le pidió a Minitab seleccionar 1000 muestras diferentes, cada una compuesta de $n = 15$ observaciones y calcular el valor de la media muestral $\overline{X}$ para cada una. La figura 5.17 es un histograma de los 1000 valores resultantes; ésta es la distribución aproximada de *x* de la toma de muestras en las circunstancias especificadas. Está claro que esta distribución es aproximadamente normal aunque el tamaño de la muestra es en realidad mucho

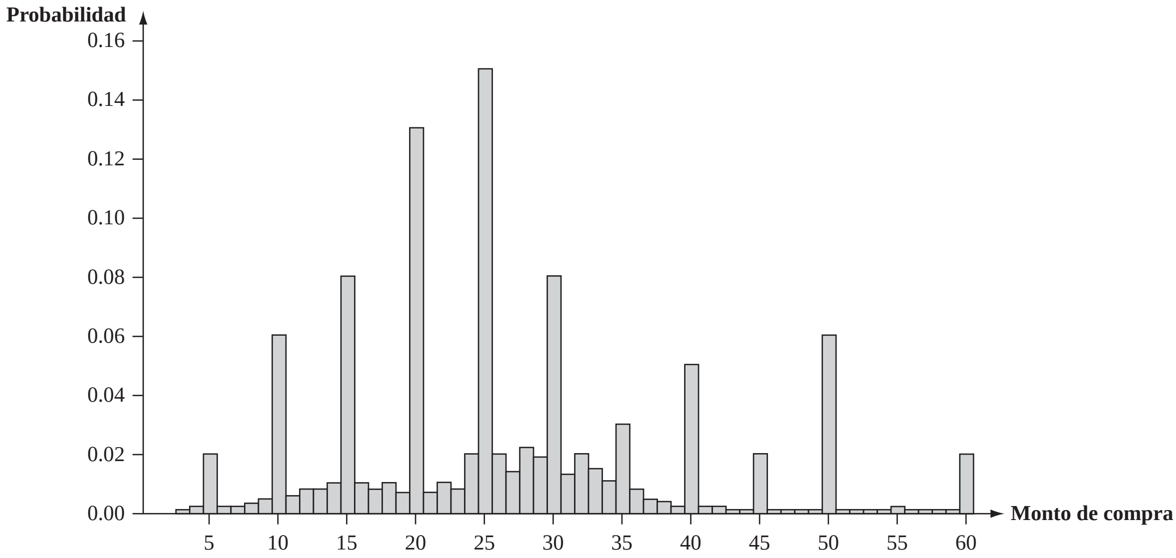


Figura 5.16 Distribución de probabilidad de X = cantidad de gasolina comprada ($)

más pequeña que 30, nuestra regla de oro de corte para invocar el teorema del límite central. Como una prueba más de la normalidad, la figura 5.18 muestra una gráfica de probabilidad normal de los 1000 valores de $\bar{x}$; el patrón lineal es muy prominente. Generalmente no es la no-normalidad en la parte central de la distribución de la población lo que hace que el TLC falle, sino una asimetría muy importante.

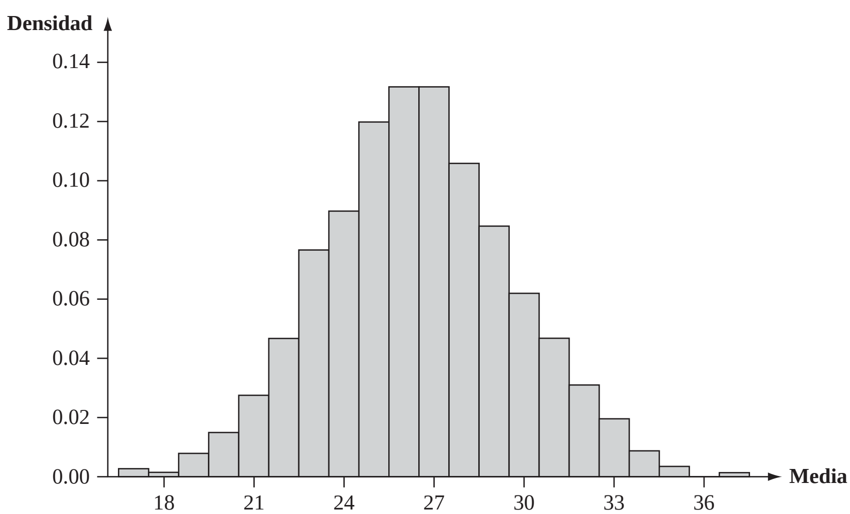


Figura 5.17 Distribución de muestreo aproximada de la media muestral de la cantidad comprada cuando n = 15 y la distribución de población es como se ve en la figura 5.16

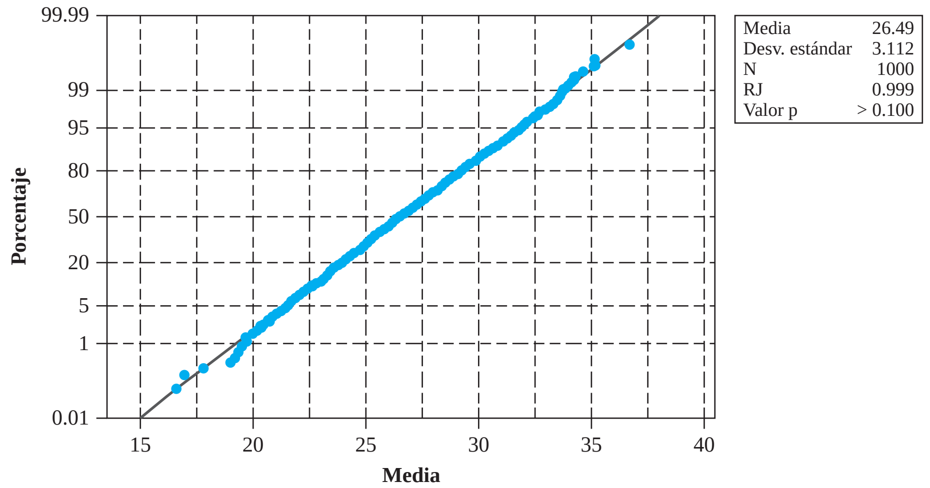


Figura 5.18 Gráfica de la probabilidad normal obtenida con Minitab de los 1000 valores de $\bar{x}$ basada en muestras de tamaño $n = 15$ ■

Otras aplicaciones del teorema del límite central

El teorema del límite central puede ser utilizado para justificar la aproximación normal a la distribución binomial discutida en el capítulo 4. Recuérdese que una variable binomial X es el número de éxitos en una experiencia binomial compuesta de n ensayos independientes con éxitos/fallas y $p = P(S)$ para cualquier ensayo particular. Defina una nueva variable aleatoria X_1 como

$$X_1 = \begin{cases} 1 & \text{si el primer ensayo produce un éxito} \\ 0 & \text{si el primer ensayo produce una falla} \end{cases}$$

y defina $X_2, X_3, \ldots, X_n$ de manera análoga para los otros $n - 1$ ensayos. Cada X_i indica si existe o no un éxito en el ensayo correspondiente.

Como los ensayos son independientes y $P(S)$ es constante de un ensayo a otro, las X_i son independientes e idénticamente distribuidas (una muestra aleatoria de una distribución de Bernoulli). El teorema del límite central implica entonces que si n es suficientemente grande, tanto la suma como el promedio de las X_i tienen distribuciones normales de manera aproximada. Cuando se suman las X_i, se agrega un 1 por cada S (éxito) que ocurra y un 0 por cada F (falla), por tanto $X_1 + \cdots + X_n = X$. La media muestral de las X_i es X/n, la proporción muestral de éxitos. Es decir, tanto X como X/n son aproximadamente normales cuando n es grande. El tamaño de muestra necesario para esta aproximación depende del valor de p: cuando p se acerca a .5, la distribución de cada X_i es razonablemente simétrica (véase la figura 5.19), mientras que la distribución es bastante asimétrica cuando p se acerca a 0 o 1. Utilizando la aproximación sólo si $np \geq 10$ y $n(1 - p) \geq 10$ garantiza que n es suficientemente grande para superar cualquier asimetría en la distribución de Bernoulli subyacente.

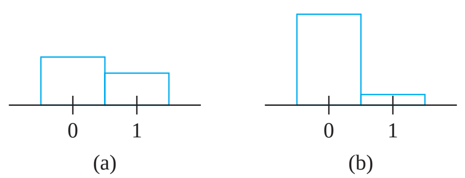


Figura 5.19 Dos distribuciones de Bernoulli: (a) $p = .4$ (razonablemente simétrica); (b) $p = .1$ (muy asimétrica)

Recuérdese de la sección 4.5 que X tiene una distribución lognormal si $\ln(X)$ tiene una distribución normal.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución para la cual sólo son posibles valores positivos $[P(X_i > 0) = 1]$. Entonces si n es suficientemente grande, el producto $Y = X_1 X_2 \cdot \cdots \cdot X_n$ tiene aproximadamente una distribución lognormal.

Para verificar esto, obsérvese que

$$\ln(Y) = \ln(X_1) + \ln(X_2) + \cdots + \ln(X_n)$$

Como $\ln(Y)$ es una suma de variables aleatorias independientes y distribuidas de manera idéntica [las $\ln(X_i)$], es aproximadamente normal cuando n es grande, así que Y tiene aproximadamente una distribución lognormal. Como ejemplo de la aplicabilidad de este resultado, Bury (*Statistical Models in Applied Science*, Wiley, pág. 590) argumenta que el proceso de daños en el flujo plástico y en la propagación de grietas es un proceso multiplicativo, de modo que variables tales como el porcentaje de alargamiento y la resistencia a la ruptura tienen aproximadamente distribuciones lognormales.

EJERCICIOS Sección 5.4 (46–57)

46. El diámetro interno de un anillo de pistón seleccionado al azar es una variable aleatoria con valor medio de 12 cm y desviación estándar de .04 cm.

a. Si $\overline{X}$ es el diámetro medio en una muestra aleatoria de $n = 16$ anillos, ¿dónde está centrada la distribución muestral de $\overline{X}$ y cuál es la desviación estándar de la distribución $\overline{X}$?

b. Responda las preguntas planteadas en el inciso (a) con un tamaño de muestra de $n = 64$ anillos.

c. ¿Con cuál de las dos muestras aleatorias, la del inciso (a) o la del inciso (b), es más probable que $\overline{X}$ esté dentro de .01 cm de 12 cm? Explique su razonamiento.

47. Remítase al ejercicio 46. Suponga que la distribución del diámetro es normal.

a. Calcule $P(11.99 \le \overline{X} \le 12.01)$, cuando $n = 16$.

b. ¿Qué tan probable es que el diámetro medio muestral exceda de 12.01 cuando $n = 25$?

48. La National Health Statistics Reports en un informe de fecha 22 de octubre de 2008, declaró que para un tamaño de muestra de 277 hombres estadounidenses de 18 años de edad, la media muestral de la circunferencia de la cintura fue de 86.3 cm. Un método algo complicado se utilizó para *estimar* varios percentiles de la población, dando como resultado los siguientes valores:

5º	10º	25º	50º	75º	90º	95º
69.6	70.9	75.2	81.3	95.4	107.1	116.4

a. ¿Es plausible que la distribución de tamaño de la cintura sea por lo menos aproximadamente normal? Explique su razonamiento. Si su respuesta es no, haga una conjetura de la forma de la distribución de la población.

b. Supongamos que la media poblacional del tamaño de la cintura es de 85 cm y que la desviación estándar de la población es de 15 cm. ¿Qué tan probable es que una muestra aleatoria de 277 individuos resulte en una media muestral del tamaño de la cintura de al menos 86.3 cm?

c. Volviendo al inciso (b), supongamos ahora que la media poblacional del tamaño de la cintura es de 82 cm. Ahora, ¿cuál es la probabilidad (aproximada) de que la media de la muestra será de al menos 86.3 cm? A la luz de este cálculo, ¿cree usted que 82 cm es un valor razonable para μ?

49. Hay 40 estudiantes en una clase de estadística elemental. Basado en años de experiencia, el instructor sabe que el tiempo requerido para calificar un primer examen seleccionado al azar es una variable aleatoria con un valor esperado de 6 min y una desviación estándar de 6 min.

a. Si los tiempos de calificación son independientes y el instructor comienza a calificar a las 6:50 p.m., y califica en forma continua, ¿cuál es la probabilidad (aproximada) de que termine de calificar antes de que se inicie el programa de noticias de las 11:00 p.m.?

b. Si el reporte de deportes se inicia a la 11:10, ¿cuál es la probabilidad de que se pierda una parte del reporte si espera hasta que termine de calificar para prender la TV?

50. La resistencia a la ruptura de un remache tiene un valor medio de 10,000 lb/pulg^2 y una desviación estándar de 500 lb/pulg^2.

a. ¿Cuál es la probabilidad de que la resistencia a la ruptura media de una muestra aleatoria de 40 remaches sea de entre 9900 y 10,200?

b. Si el tamaño de muestra hubiera sido de 15 y no de 40, ¿se podría calcular la probabilidad solicitada en el inciso (a) con la información dada?

51. El tiempo requerido por un solicitante de una hipoteca seleccionado al azar para llenar un formulario tiene una distribución normal con valor medio de 10 minutos y desviación estándar de 2 min. Si cinco individuos llenan un formulario en un día y seis en otro, ¿cuál es la probabilidad de que la cantidad de tiempo promedio muestral requerido cada día sea cuando mucho de 11 minutos?

52. La vida útil de un tipo de batería está normalmente distribuida con valor medio de 10 horas y desviación estándar de 1 hora.

Hay cuatro baterías en un paquete. ¿Qué valor de vida útil es tal que la vida útil total de todas las baterías contenidas en un paquete exceda ese valor en sólo 5% de todos los paquetes?

53. Se sabe que la dureza Rockwell de pines de un tipo tiene un valor medio de 50 y una desviación estándar de 1.2.
   a. Si la distribución es normal, ¿cuál es la probabilidad de que la dureza media de una muestra aleatoria de 9 pines sea por lo menos de 51?
   b. Sin suponer una población normal, ¿cuál es la probabilidad (aproximada) de que la dureza media de una muestra aleatoria de 40 pines sea por lo menos de 51?

54. Suponga que la densidad de un sedimento (g/cm) de un espécimen seleccionado al azar de cierta región está normalmente distribuida con media de 2.65 y desviación estándar de .85 (sugerida en "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants", *Water Research*, 1984: 1169–1174).
   a. Si se selecciona una muestra aleatoria de 25 especímenes, ¿cuál es la probabilidad de que la densidad del sedimento promedio muestral sea cuando mucho de 3.00? ¿De entre 2.65 y 3.00?
   b. ¿Qué tan grande debe ser un tamaño de muestra para garantizar que la primera probabilidad en el inciso (a) sea por lo menos de .99?

55. El número de infracciones de estacionamiento aplicadas en una ciudad en cualquier día dado de la semana tiene una distribución de Poisson con parámetro $\mu = 50$. ¿Cuál es la probabilidad aproximada de que
   a. entre 35 y 70 infracciones sean aplicadas en un día particular? [*Sugerencia*: cuando μ es grande, una variable aleatoria de Poisson tiene aproximadamente una distribución normal.]
   b. el número total de infracciones aplicadas durante una semana de 5 días sea de entre 225 y 275?

56. Un canal de comunicación binaria transmite una secuencia de "bits" (ceros y unos). Suponga que por cualquier bit particular transmitido, existe 10% de probabilidad de que ocurra un error en la transmisión (un 0 se convierte en 1 o un 1 se convierte en 0). Suponga que los errores en los bits ocurren independientemente uno de otro.
   a. Considere transmitir 1000 bits. ¿Cuál es la probabilidad aproximada de que cuando mucho ocurran 125 errores de transmisión?
   b. Suponga que el mismo mensaje de 1000 bits es enviado en dos momentos diferentes independientemente uno de otro. ¿Cuál es la probabilidad aproximada de que el número de errores en la primera transmisión esté dentro de 50 del número de errores en la segunda?

57. Suponga que la distribución del tiempo X (en horas) utilizado por estudiantes en cierta universidad en un proyecto particular es gamma con parámetros $\alpha = 50$ y $\beta = 2$. Como α es grande, se puede demostrar que X tiene aproximadamente una distribución normal. Use este hecho para calcular la probabilidad aproximada de que un estudiante seleccionado al azar utilice cuando mucho 125 horas en el proyecto.

5.5 Distribución de una combinación lineal

La media muestral $\overline{X}$ y el total muestral T_o son casos especiales de un tipo de variable aleatoria que surgen con frecuencia en aplicaciones estadísticas.

DEFINICIÓN

Dado un conjunto de n variables aleatorias $X_1, \ldots, X_n$ y n constantes numéricas $a_1, \ldots, a_n$, la variable aleatoria

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \tag{5.7}$$

se llama **combinación lineal** de las X_i.

Por ejemplo, $4X_1 - 5X_2 + 8X_3$ es una combinación lineal de X_1, X_2 y X_3 con $a_1 = 4$; $a_2 = -5$ y $a_3 = 8$.

Tomando $a_1 = a_2 = \cdots = a_n = 1$ da como resultado $Y = X_1 + \cdots + X_n = T_o$ y $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$ resulta en

$$Y = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}T_o = \overline{X}$$

Obsérvese que no se requiere que las X_i sean independientes o estén idénticamente distribuidas. Todas las X_i podrían tener distribuciones diferentes y por consiguiente valores medios y varianzas diferentes. Primero se considera el valor esperado y la varianza de una combinación lineal.

PROPOSICIÓN

Sean $X_1, X_2, \ldots, X_n$ con valores medios $\mu_1, \ldots, \mu_n$, respectivamente, y varianzas $\sigma_1^2, \ldots, \sigma_n^2$, respectivamente.

1. Si las X_i son independientes o no

$$\begin{aligned} E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\ &= a_1\mu_1 + \cdots + a_n\mu_n \end{aligned} \tag{5.8}$$

2. Si $X_1, \ldots, X_n$ son independientes,

$$\begin{aligned} V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\ &= a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \end{aligned} \tag{5.9}$$

y

$$\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2} \tag{5.10}$$

3. Con cualquier $X_1, \ldots, X_n$,

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\mathrm{Cov}(X_i, X_j) \tag{5.11}$$

Las comprobaciones se dan al final de la sección. Un parafraseo de (5.8) es que el valor esperado de una combinación lineal es la misma combinación lineal de los valores esperados; por ejemplo, $E(2X_1 + 5X_2) = 2\mu_1 + 5\mu_2$. El resultado (5.9) en la proposición 2 es un caso especial de (5.11) en la proposición 3; cuando las X_i son independientes, $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$ y $= V(X_i)$ para $i = j$ (esta simplificación en realidad ocurre cuando las X_i no están correlacionadas, una condición más débil que la de independencia). Especializando al caso de una muestra aleatoria (X_i independientes e idénticamente distribuidas) con $a_i = 1/n$ para cada i da $E(\overline{X}) = \mu$ y $V(\overline{X}) = \sigma^2/n$ como se discutió en la sección 5.4. Un comentario similar se aplica a las reglas para T_o.

Ejemplo 5.29 Una gasolinera vende tres grados de gasolina; regular, extra y súper. Éstas se venden a \$3.00, \$3.20 y \$3.40 por galón, respectivamente. Sean X_1, X_2 y X_3 las cantidades (galones) de estas gasolinas compradas en un día particular. Suponga que las X_i son independientes con $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$ y $\sigma_3 = 50$. El ingreso por las ventas es $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$ y

$$\begin{aligned} E(Y) &= 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620 \\ V(Y) &= (3.0)^2\sigma_1^2 + (3.2)^2\sigma_2^2 + (3.4)^2\sigma_3^2 = 184{,}436 \\ \sigma_Y &= \sqrt{184{,}436} = \$429.46 \end{aligned}$$

■

Diferencia entre dos variables aleatorias

Un importante caso especial de una combinación lineal se presenta con $n = 2$, $a_1 = 1$ y $a_2 = -1$:

$$Y = a_1X_1 + a_2X_2 = X_1 - X_2$$

Entonces se tiene el siguiente corolario de la proposición.

COROLARIO

$E(X_1 - X_2) = E(X_1) - E(X_2)$ para dos variables aleatorias cualesquiera X_1 y X_2. $V(X_1 - X_2) = V(X_1) + V(X_2)$ si X_1 y X_2 son variables aleatorias independientes.

El valor esperado de una diferencia es la diferencia de los dos valores esperados, pero la varianza de una diferencia entre dos variables independientes es la *suma*, *no* la diferencia, de las dos varianzas. Existe tanta variabilidad en $X_1 - X_2$ como en $X_1 + X_2$ [escribiendo $X_1 - X_2 = X_1 + (-1)X_2$; $(-1)X_2$ tiene la misma cantidad de variabilidad que X_2].

Ejemplo 5.30 Una compañía automotriz equipa un modelo particular con un motor de seis cilindros o un motor de cuatro cilindros. Sean X_1 y X_2 eficiencias de combustible de automóviles de seis y cuatro cilindros seleccionados en forma independiente al azar, respectivamente. Con $\mu_1 = 22$, $\mu_2 = 26$, $\sigma_1 = 1.2$ y $\sigma_2 = 1.5$,

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$
$$V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 = (1.2)^2 + (1.5)^2 = 3.69$$
$$\sigma_{X_1 - X_2} = \sqrt{3.69} = 1.92$$

Si se cambia la notación de modo que X_1 se refiera al automóvil de cuatro cilindros, entonces $E(X_1 - X_2) = 4$, pero la varianza de la diferencia sigue siendo de 3.69. ■

El caso de variables aleatorias normales

Cuando las X_i forman una muestra aleatoria de una distribución normal, $\overline{X}$ y T_o están normalmente distribuidas. He aquí un resultado más general con respecto a combinaciones lineales.

PROPOSICIÓN

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes normalmente distribuidas (con quizá diferentes medias y/o varianzas), entonces cualquier combinación lineal de las X_i también tiene una distribución normal. En particular, la diferencia $X_1 - X_2$ entre dos variables independientes normalmente distribuidas también está distribuida en forma normal.

Ejemplo 5.31 (Continuación del ejemplo 5.29) El ingreso total por la venta de los tres grados de gasolina en un día particular fue $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$ y se calculó $\mu_Y = 5620$ y (suponiendo independencia) $\sigma_Y = 429.46$. Si las X_i están normalmente distribuidas, la probabilidad de que el ingreso sea de más de 4500 es

$$P(Y > 4500) = P\left(Z > \frac{4500 - 5620}{429.46}\right)$$
$$= P(Z > -2.61) = 1 - \Phi(-2.61) = .9955$$ ■

El teorema del límite central también puede ser generalizado para aplicarlo a ciertas combinaciones lineales. En general, si n es grande y no es probable que algún término individual contribuya demasiado al valor total, entonces Y tiene aproximadamente una distribución normal.

Comprobaciones en el caso $n = 2$

En cuanto al resultado por lo concerniente a los valores esperados, suponga que X_1 y X_2 son continuas con función de densidad de probabilidad conjunta $f(x_1, x_2)$. Entonces

$$E(a_1X_1 + a_2X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(a_1x_1 + a_2x_2)f(x_1, x_2)\,dx_1\,dx_2$$
$$= a_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1, x_2)\,dx_2\,dx_1$$
$$+ a_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2f(x_1, x_2)\,dx_1\,dx_2$$
$$= a_1\int_{-\infty}^{\infty}x_1f_{X_1}(x_1)\,dx_1 + a_2\int_{-\infty}^{\infty}x_2f_{X_2}(x_2)\,dx_2$$
$$= a_1E(X_1) + a_2E(X_2)$$

La suma reemplaza a la integración en el caso discreto. El argumento en cuanto a la varianza resultante no requiere especificar si la variable es discreta o continua. Recordando que $V(Y) = E[(Y - \mu_Y)^2]$,

$$V(a_1X_1 + a_2X_2) = E\{[a_1X_1 + a_2X_2 - (a_1\mu_1 + a_2\mu_2)]^2\}$$
$$= E\{a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\}$$

La expresión dentro de los paréntesis rectangulares es una combinación lineal de las variables $Y_1 = (X_1 - \mu_1)^2$, $Y_2 = (X_2 - \mu_2)^2$ y $Y_3 = (X_1 - \mu_1)(X_2 - \mu_2)$, así que si se acarrea la operación E a través de los tres términos se obtiene $a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2\,\text{Cov}(X_1, X_2)$ como se requiere. ■

EJERCICIOS Sección 5.5 (58–74)

58. Una compañía naviera maneja contenedores en tres diferentes tamaños: (1) 27 pies^3 ($3 \times 3 \times 3$), (2) 125 pies^3 y (3) 512 pies^3. Sea X_i ($i = 1, 2, 3$) el número de contenedores de tipo i embarcados durante una semana dada. Con $\mu_i = E(X_i)$ y $\sigma_i^2 = V(X_i)$, suponga que los valores medios y las desviaciones estándar son como sigue.

$\mu_1 = 200$	$\mu_2 = 250$	$\mu_3 = 100$
$\sigma_1 = 10$	$\sigma_2 = 12$	$\sigma_3 = 8$

a. Suponiendo que X_1, X_2, X_3 son independientes, calcule el valor esperado y la varianza del volumen total embarcado. [*Sugerencia*: volumen $= 27X_1 + 125X_2 + 512X_3$.]

b. ¿Serían sus cálculos necesariamente correctos si las X_i no fueran independientes? Explique.

59. Sean X_1, X_2 y X_3 que representan los tiempos necesarios para realizar tres tareas de reparación sucesivas en cierto taller de servicio. Suponga que son variables aleatorias normales independientes con valores esperados μ_1, μ_2 y μ_3 y varianzas σ_1^2, σ_2^2 y σ_3^2, respectivamente.

a. Si $\mu_1 = \mu_2 = \mu_3 = 60$ y $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 15$ calcule $P(T_o \le 200)$. ¿Cuál es $P(150 \le T_o \le 200)$?

b. Con las μ_i y σ_i dadas en el inciso (a), calcule $P(55 \le \overline{X})$ y $P(58 \le \overline{X} \le 62)$.

c. Con las μ_i y σ_i dadas en el inciso (a), calcule e interprete $P(-10 \le X_1 - .5X_2 - .5X_3 \le 5)$.

d. Si $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ y $\sigma_3^2 = 14$, calcule $P(X_1 + X_2 + X_3 \le 160)$ y $P(X_1 + X_2 \ge 2X_3)$.

60. Cinco automóviles del mismo tipo tienen que realizar un viaje de 300 millas. Los primeros dos utilizarán una marca económica de gasolina y los otros tres una marca de renombre. Sean X_1, X_2, X_3, X_4 y X_5 las eficiencias de combustible observadas (mpg) de los cinco carros. Suponga que estas variables son independientes y normalmente distribuidas con $\mu_1 = \mu_2 = 20$, $\mu_3 = \mu_4 = \mu_5 = 21$ y $\sigma^2 = 4$ con la marca económica y 3.5 con la marca de renombre. Defina una variable aleatoria Y como

$$Y = \frac{X_1 + X_2}{2} - \frac{X_3 + X_4 + X_5}{3}$$

de modo que Y mide la diferencia de eficiencia entre la gasolina económica y la de renombre. Calcule $P(0 \le Y)$ y $P(-1 \le Y \le 1)$. [*Sugerencia*: $Y = a_1X_1 + \cdots + a_5X_5$, con $a_1 = \frac{1}{2}, \ldots, a_5 = -\frac{1}{3}$.]

61. El ejercicio 26 introdujo variables aleatorias X y Y, el número de carros y autobuses, respectivamente, transportados por un transbordador en un solo viaje. La función de masa de probabilidad conjunta de X y Y se da en la tabla del ejercicio 7. Es fácil verificar que X y Y son independientes.

a. Calcule el valor esperado, la varianza y la desviación estándar del número total de vehículos en un solo viaje.

b. Si a cada carro se le cobran \$3 y a cada autobús \$10, calcule el valor esperado, la varianza y la desviación estándar del ingreso resultante de un solo viaje.

62. Un fabricante de cierto componente requiere tres operaciones de maquinado diferentes. El tiempo de maquinado de cada operación tiene una distribución normal y los tres tiempos son independientes entre sí. Los valores medios son 15, 30 y 20 min, respectivamente, y las desviaciones estándar son 1, 2 y 1.5 min, respectivamente. ¿Cuál es la probabilidad de que se requiera cuando mucho 1 hora de tiempo de maquinado para producir un componente seleccionado al azar?

63. Remítase al ejercicio 3.

a. Calcule la covarianza entre X_1 = el número de clientes en la caja rápida y X_2 = el número de clientes en la caja extrarrápida.

b. Calcule $V(X_1 + X_2)$. ¿Cómo se compara con $V(X_1) + V(X_2)$?

64. Suponga que el tiempo de espera para un autobús en la mañana está uniformemente distribuido en [0, 8], mientras que el tiempo de espera en la noche está uniformemente distribuido en [0, 10] independiente del tiempo de espera en la mañana.

a. Si toma el autobús en la mañana y en la noche durante una semana, ¿cuál es su tiempo de espera total esperado? [*Sugerencia*: defina las variables aleatorias $X_1, \ldots, X_{10}$ y use una regla de valor esperado.]

b. ¿Cuál es la varianza de su tiempo de espera total?

c. ¿Cuáles son el valor esperado y la varianza de la diferencia entre los tiempos de espera en la mañana y en la noche en un día dado?

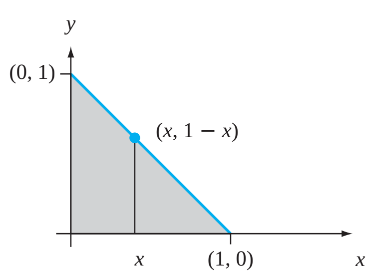
d. ¿Cuáles son el valor esperado y la varianza de la diferencia entre el tiempo de espera total en la mañana y el tiempo de espera total en la noche durante una semana particular?

65. Suponga que cuando el pH de cierto compuesto químico es 5.00, el pH medido por un estudiante de química de primer año seleccionado al azar es una variable aleatoria con media de 5.00 y desviación estándar de .2. Un gran lote del compuesto se subdivide y a cada estudiante se le da una muestra en un laboratorio matutino y a cada estudiante en un laboratorio vespertino. Sea $\overline{X}$ = el pH promedio determinado por los estudiantes matutinos y $\overline{Y}$ = el pH promedio determinado por los estudiantes vespertinos.

a. Si el pH es una variable normal y hay 25 estudiantes en cada laboratorio, calcule $P(-.1 \le \overline{X} - \overline{Y} \le .1)$. [*Sugerencia*: $\overline{X} - \overline{Y}$ es una combinación lineal de variables normales, así que está normalmente distribuida. Calcule $\mu_{\overline{X}-\overline{Y}}$ y $\sigma_{\overline{X}-\overline{Y}}$.]

b. Si hay 36 estudiantes en cada laboratorio, pero las determinaciones del pH no se suponen normales, calcule (aproximadamente) $P(-.1 \le \overline{X} - \overline{Y} \le .1)$.

66. Si se aplican dos cargas a una viga en voladizo como se muestra en la figura adjunta, el momento de flexión en 0 debido a las cargas es $a_1X_1 + a_2X_2$.



a. Suponga que X_1 y X_2 son variables aleatorias independientes con medias de 2 y 4 kips, respectivamente, y desviaciones estándar de .5 y 1.0 kip, respectivamente. Si $a_1 = 5$ pies y $a_2 = 10$ pies, ¿cuál es el momento de flexión esperado y cuál es la desviación estándar del momento de flexión?

b. Si X_1 y X_2 están normalmente distribuidas, ¿cuál es la probabilidad de que el momento de flexión sea de más de 75 kips-pie?

c. Suponga que las posiciones de las dos cargas son variables aleatorias. Denotándolas por A_1 y A_2, suponga que estas variables tienen medias de 5 y 10 pies, respectivamente, que cada una tiene una desviación estándar de .5, y que todas las A_i y X_i son independientes entre sí. ¿Cuál es el momento esperado ahora?

d. En la situación del inciso (c), ¿cuál es la varianza del momento de flexión?

e. Si la situación es como se describe en el inciso (a) excepto que $\text{Corr}(X_1, X_2) = .5$ (de modo que las dos cargas no sean independientes), ¿cuál es la varianza del momento de flexión?

67. Un tramo de tubería de PVC tiene que ser insertado en otro tramo. La longitud del primer tramo está normalmente distribuida con valor medio de 20 pulg y desviación estándar de .5 pulg. La longitud del segundo tramo es una variable aleatoria normal con media y desviación estándar de 15 pulg y .4 pulg, respectivamente. La cantidad de traslape está normalmente distribuida con valor medio de 1 pulg y desviación estándar de .1 pulg. Suponiendo que los tramos y cantidad de traslape son independientes entre sí, ¿cuál es la probabilidad de que la longitud total después de la inserción sea de entre 34.5 pulg y 35 pulg?

68. Dos aviones vuelan en la misma dirección en dos corredores paralelos adyacentes. En el instante $t = 0$, el primer avión está a 10 km adelante del segundo. Suponga que la velocidad del primer avión (km/h) está normalmente distribuida con media de 520 y desviación estándar de 10 y que la velocidad del segundo también está normalmente distribuida con media y desviación estándar de 500 y 10, respectivamente.

a. ¿Cuál es la probabilidad de que después de 2 horas de vuelo el segundo avión no haya alcanzado al primer avión?

b. Determine la probabilidad de que los aviones estén separados cuando mucho 10 km después de 2 horas.

69. Tres carreteras diferentes entroncan en la entrada de una autopista particular. Suponga que durante un tiempo fijo, el número de carros que llegan por cada carretera a la autopista es una variable aleatoria con valor esperado y desviación estándar como se dan en la tabla

	Carretera 1	Carretera 2	Carretera 3
Valor esperado	800	1000	600
Desviación estándar	16	25	18

a. ¿Cuál es el número de carros total esperado que entran a la autopista en este punto durante el periodo? [*Sugerencia*: sea X_i = el número de la carretera i.]

b. ¿Cuál es la varianza del número total de carros que entran? ¿Ha hecho suposiciones sobre la relación entre los números de carros en las diferentes carreteras?

c. Con X_i denotando el número de carros que entran de la carretera i durante el periodo, suponga que $\text{Cov}(X_1, X_2) = 80$, $\text{Cov}(X_1, X_3) = 90$, y $\text{Cov}(X_2, X_3) = 100$ (de modo que las tres corrientes de tráfico no son independientes). Calcule el número total esperado de los carros que entran y la desviación estándar del total.

70. Considere una muestra aleatoria de tamaño n tomada de una distribución continua con mediana 0 de modo que la probabilidad de que cualquier observación sea positiva es de .5. Haciendo caso omiso de los signos de las observaciones, clasifíquelas desde las más pequeña a la más grande en valor absoluto y sea W = la suma de los renglones de las observaciones con signos positivos. Por ejemplo, si las observaciones son $-.3$, $+.7$, $+2.1$ y -2.5, entonces los renglones de observaciones positivas son 2 y 3, de modo que $W = 5$. En el capítulo 15, W se llamará *estadístico de filas con signo de Wilcoxon*. W puede representarse como sigue:

$$W = 1 \cdot Y_1 + 2 \cdot Y_2 + 3 \cdot Y_3 + \cdots + n \cdot Y_n$$
$$= \sum_{i=1}^{n} i \cdot Y_i$$

donde las Y_i son variables aleatorias independientes de Bernoulli, cada una con $p = .5$ ($Y_i = 1$ corresponde a la observación con fila i positiva).

a. Determine $E(Y_i)$ y luego $E(W)$ utilizando la ecuación para W. [*Sugerencia*: los primeros n enteros positivos se suman a $n(n + 1)/2$.]

b. Determine $V(Y_i)$ y luego $V(W)$. [*Sugerencia*: la suma de los cuadrados de los primeros n enteros positivos puede expresarse como $n(n + 1)(2n + 1)/6$.]

71. En el ejercicio 66, el peso de la viga contribuye al momento de flexión. Suponga que la viga es de espesor y densidad uniformes, de modo que la carga resultante esté uniformemente distribuida en la viga. Si el peso de ésta es aleatorio, la carga resultante a consecuencia del peso también es aleatoria; denote esta carga por W (kip-pies).

a. Si la viga es de 12 pies de largo, W tiene una media de 1.5 y una desviación estándar de .25 y las cargas fijas son como se describen en el inciso (a) del ejercicio 66, ¿cuáles son el valor esperado y la varianza del momento de flexión? [*Sugerencia*: si la carga originada por la viga fuera w kip-pies, la contribución al momento de flexión sería $w\int_0^{12} x\,dx$.]

b. Si las tres variables (X_1, X_2 y W) están normalmente distribuidas, ¿cuál es la probabilidad de que el momento de flexión será cuando mucho de 200 kip-pies?

72. Tengo tres encargos que atender en el Edificio de Administración. Sea X_i = el tiempo que requiere el encargo i-ésimo (i = 1, 2, 3) y sea X_4 = el tiempo total en minutos que me paso caminando hasta el edificio y de regreso, y entre cada encargo. Suponga que las X_i son independientes y normalmente distribuidas con las siguientes medias y desviaciones estándar: $\mu_1 = 15$, $\sigma_1 = 4$, $\mu_2 = 5$, $\sigma_2 = 1$, $\mu_3 = 8$, $\sigma_3 = 2$, $\mu_4 = 12$, $\sigma_4 = 3$. Pienso salir de mi oficina precisamente a las 10:00 a.m. y deseo pegar una nota en la puerta que diga "Regreso alrededor de las t a.m.". ¿Qué hora debo escribir si deseo que la probabilidad de mi llegada después de t sea de .01?

73. Suponga que la resistencia a la tensión esperada de acero tipo A es de 105 kg/pulg^2 y que la desviación estándar de la resistencia a la tensión es de 8 kg/pulg^2. Para acero tipo B, suponga que la resistencia a la tensión esperada y la desviación estándar de la resistencia a la tensión son de 100 k/pulg^2 y 6 kg/pulg^2, respectivamente. Sea $\overline{X}$ = la resistencia a la tensión promedio de una muestra aleatoria de 40 especímenes tipo A, y sea $\overline{Y}$ = la resistencia a la tensión promedio de una muestra aleatoria de 35 especímenes tipo B.

a. ¿Cuál es la distribución aproximada de $\overline{X}$? ¿De $\overline{Y}$?

b. ¿Cuál es la distribución aproximada de $\overline{X} - \overline{Y}$? Justifique su respuesta.

c. Calcule (aproximadamente) $P(-1 \leq \overline{X} - \overline{Y} \leq 1)$.

d. Calcule $P(\overline{X} - \overline{Y} \geq 10)$. Si realmente observa que $\overline{X} - \overline{Y} \geq 10$, ¿le cabe duda de que $\mu_1 - \mu_2 = 5$?

74. En un área de suelo arenoso se plantaron 50 árboles pequeños de un cierto tipo, y otros 50 se plantaron en un área de suelo arcilloso. Sea X = el número de árboles plantados en suelo arenoso que sobreviven 1 año y Y = el número de árboles plantados en suelo arcilloso que sobreviven 1 año. Si la probabilidad de que un árbol plantado en suelo arenoso sobreviva 1 año es de .7 y la probabilidad de sobrevivencia de 1 año en suelo arcilloso es de .6, calcule una aproximación a $P(-5 \leq X - Y \leq 5)$ (no se moleste con la corrección de continuidad).

EJERCICIOS SUPLEMENTARIOS (75-96)

75. Un restaurante sirve tres comidas que cuestan \$12, \$15 y \$20. Para una pareja seleccionada al azar que está comiendo en este restaurante, sea X = el costo de la comida del hombre y Y = el costo de la comida de la mujer. La función de masa de probabilidad conjunta de X y Y se da en la siguiente tabla:

$p(x, y)$		y = 12	15	20
	12	.05	.05	.10
x	15	.05	.10	.35
	20	0	.20	.10

a. Calcule las funciones de masa de probabilidad marginal de X y Y.

b. ¿Cuál es la probabilidad de que las comidas del hombre y la mujer cuesten cuando mucho \$15 cada una?

c. ¿Son X y Y independientes? Justifique su respuesta.

d. ¿Cuál es el costo total esperado de la comida de las dos personas?

e. Suponga que cuando una pareja abre las galletas de la fortuna al final de la comida, encuentran el mensaje: "Recibirá como reembolso la diferencia entre el costo de la comida más cara y la menos cara que eligió". ¿Cuánto espera reembolsar el restaurante?

76. En una estimación de costos, el costo total de un proyecto es la suma de los costos de las tareas componentes. Cada uno de estos costos es una variable aleatoria con una distribución de probabilidad. Se acostumbra obtener información sobre la distribución de costos total sumando las características de las distribuciones de costo de componente individuales; esto se conoce como procedimiento de "despliegue". Por ejemplo, $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$, así que el procedimiento de despliegue es válido para costo medio. Suponga que hay dos tareas componentes y que X_1 y X_2 son variables aleatorias independientes normalmente distribuidas. ¿Es válido el procedimiento de despliegue para el 75º percentil? Es decir, ¿Es el 75º percentil de la distribución de $X_1 + X_2$ el mismo que la suma de los 75º percentiles de las dos distribuciones individuales? Si no, ¿cuál es la relación entre el percentil de la suma y la suma de los percentiles? ¿Con qué percentiles es válido el procedimiento de despliegue en este caso?

77. Una tienda de comida saludable vende dos marcas diferentes de un tipo de grano. Sea X = la cantidad (lb) de la marca A disponible y Y = la cantidad de la marca B disponible. Suponga que la función de densidad de probabilidad conjunta de X y Y es

$$f(x, y) = \begin{cases} kxy & x \geq 0, y \geq 0, 20 \leq x + y \leq 30 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la región de densidad positiva y determine el valor de k.

b. ¿Son X y Y independientes? Responda obteniendo primero la función de densidad de probabilidad marginal de cada variable.

c. Calcule $P(X + Y \leq 25)$.

d. ¿Cuál es la cantidad total esperada de este grano disponible?

e. Calcule $\text{Cov}(X, Y)$ y $\text{Corr}(X, Y)$.

f. ¿Cuál es la varianza de la cantidad total de grano disponible?

78. El artículo "Stochastic Modeling for Pavement Warranty Cost Estimation" (*J. of Constr. Engr. and Mgmnt.*, 2009: 352−359) propone el siguiente modelo para la distribución de Y = tiempo de falla para pavimento. Sean X_1 el momento de un fallo debido a la formación de roderas y sea X_2 el momento de un fallo debido a agrietamiento transversal; estas dos variables aleatorias se suponen independientes. Entonces $Y = \min(X_1, X_2)$. La probabilidad de un fallo debido a cualquiera de estos modos de alteración se supone que es una función creciente del tiempo t. Después de hacer algunas suposiciones de distribución, se obtiene la siguiente forma de la función de distribución acumulativa para cada modo:

$$\Phi\left[(a + bt) / (c + dt + et^2)^{1/2}\right]$$

donde Φ es la función de distribución acumulativa normal estándar. Los valores de los cinco parámetros a, b, c, d y e son −25.49, 1.15, 4.45, −1.78 y .171 para el agrietamiento, y −21.27, .0325, .972, −.00028 y .00022 para las roderas. Determine la probabilidad de falla de los pavimentos en $t = 5$ años y $t = 10$ años.

79. Suponga que para un individuo, la ingesta de calorías en el desayuno es una variable aleatoria con valor esperado de 500 y desviación estándar de 50, la ingesta de calorías en el almuerzo es aleatoria con valor esperado de 900 y desviación estándar de 100, y la ingesta de calorías en la comida es una variable aleatoria con valor esperado de 2000 y desviación estándar de 180. Suponiendo que las ingestas en las diferentes comidas son independientes entre sí, ¿cuál es la probabilidad de que la ingesta de calorías promedio por día durante el siguiente año (365 días) sea cuando mucho de 3500? [*Sugerencia*: sean X_i, Y_i y Z_i las tres ingestas de calorías en el día i. Entonces la ingesta total es $\Sigma(X_i + Y_i + Z_i)$.]

80. El peso medio del equipaje documentado por un pasajero de clase turista seleccionado al azar que vuela entre dos ciudades en cierta aerolínea es de 40 lb y la desviación estándar es de 10 lb. La media y la desviación estándar de un pasajero de clase de negocios son 30 lb y 6 lb, respectivamente.

a. Si hay 12 pasajeros de clase de negocios y 50 de clase turista en un vuelo particular, ¿cuáles son el valor esperado y la desviación estándar del peso total del equipaje?

b. Si los pesos individuales de los equipajes son variables aleatorias independientes normalmente distribuidas, ¿cuál es la probabilidad de que el peso total del equipaje sea cuando mucho de 2500 lb?

81. Se ha visto que si $E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$, entonces $E(X_1 + \cdots + X_n) = n\mu$. En algunas aplicaciones, el número de X_i consideradas no es un número fijo n sino una variable aleatoria N. Por ejemplo, sea N = el número de componentes que son traídos a un taller de reparación en un día particular y sea X_i el tiempo de reparación del componente i-ésimo. Entonces el tiempo de reparación total es $X_1 + X_2 + \cdots + X_N$, la suma de un número *aleatorio* de variables aleatorias. Cuando N es independiente de las X_i se puede demostrar que

$$E(X_1 + \cdots + X_N) = E(N) \cdot \mu$$

a. Si el número esperado de componentes traídos en un día particular es 10 y el tiempo de reparación esperado de un componente seleccionado al azar es de 40 min, ¿cuál es el tiempo de reparación total esperado de componentes entregados en cualquier día particular?

b. Suponga que componentes de un tipo llegan para ser reparados de acuerdo con un proceso de Poisson a razón de 5 por hora. El número esperado de defectos por componente es de 3.5. ¿Cuál es el valor esperado del número total de defectos en componentes traídos a reparación durante un periodo de 4 horas? Asegúrese de indicar cómo su respuesta se deriva del resultado general que se acaba de dar.

82. Suponga que la proporción de votantes rurales en un estado que favorecen a un candidato a gobernador particular es de .45, y que la proporción de votantes suburbanos y urbanos que favorecen al candidato es de .60. Si se obtiene una muestra de 200 votantes rurales y 300 votantes suburbanos y urbanos, ¿cuál es la probabilidad aproximada de que por lo menos 250 de estos votantes favorezcan a este candidato?

83. Sea μ el pH verdadero de un compuesto químico. Se realizará una secuencia de n determinaciones de pH muestrales independientes. Suponga que cada pH muestras es una variable aleatoria con valor esperado μ y desviación estándar de .1. ¿Cuántas determinaciones se requieren si se desea que la probabilidad de que el promedio muestral esté dentro de .02 del pH verdadero sea por lo menos de .95? ¿Qué teorema justifica su cálculo de probabilidad?

84. Si la cantidad de refresco que consumo en cualquier día dado es independiente del consumo en cualquier otro día y está normalmente distribuido con $\mu = 13$ oz y $\sigma = 2$, y si en este momento tengo dos paquetes de seis botellas de 16 oz, ¿cuál es la probabilidad de que todavía tenga algo de refresco al cabo de 2 semanas (14 días)?

85. Remítase al ejercicio 58 y suponga que las X_i son independientes entre sí y que cada una tiene una distribución normal. ¿Cuál es la probabilidad de que el volumen total embarcado sea cuando mucho de 100,000 pies^3?

86. Un estudiante tiene una clase que se supone termina a las 9:00 a.m. y otra que se supone comienza a las 9:10 a.m. Suponga que el tiempo real de terminación de la clase de las 9:00 a.m. es una variable aleatoria normalmente distribuida X_1 con media de 9:02 y desviación estándar de 1.5 min y que la hora de inicio de la siguiente clase también es una variable aleatoria normalmente distribuida X_2 con media de 9:10 y desviación estándar de 1 min. Suponga que el tiempo necesario para ir de un salón de clases al otro es una variable aleatoria normalmente distribuida X_3 con media de 6 min y desviación estándar de 1 min. ¿Cuál es la probabilidad de que el estudiante llegue a la segunda clase antes de que comience? (Suponga independencia de X_1, X_2 y X_3, lo cual es razonable si el estudiante no presta atención a la hora de terminación de la primera clase.)

87. **a.** Use la fórmula general de la varianza de una combinación lineal para escribir una expresión para $V(aX + Y)$. Luego sea $a = \sigma_Y/\sigma_X$ y demuestre que $\rho \geq -1$. [*Sugerencia*: la varianza siempre es ≥ 0 y $\text{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \cdot \rho$.]

b. Considerando $V(aX - Y)$, concluya que $\rho \leq 1$.

c. Use el hecho de que $V(W) = 0$ sólo si W es una constante para demostrar que $\rho = 1$ sólo si $Y = aX + b$.

88. Suponga que una calificación oral X y una calificación cuantitativa Y de un individuo seleccionado al azar en un examen de aptitud administrado nacionalmente tienen una función de densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \dfrac{2}{5}(2x + 3y) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Se le pide que haga una predicción t de la calificación total $X + Y$ del individuo. El error de predicción es la media del error al cuadrado $E[(X + Y - t)^2]$. ¿Qué valor de t reduce al mínimo el error de predicción?

89. **a.** Sea X_i que tiene una distribución ji cuadrada con parámetro ν_1 (véase la sección 4.4) y sea X_2 independiente de X_1 que tiene una distribución ji cuadrada con parámetro ν_2. Use la técnica del ejemplo 5.21 para demostrar que $X_1 + X_2$ tiene una distribución ji cuadrada con parámetro $\nu_1 + \nu_2$.

b. En el ejercicio 71 del capítulo 4, se le pidió que demostrara que si Z es una variable aleatoria normal estándar, entonces Z^2 tiene una distribución ji cuadrada con $\nu = 1$. Sean $Z_1, Z_2, \ldots, Z_n$ n variables aleatorias normales estándar independientes. ¿Cuál es la distribución de $Z_1^2 + \cdots + Z_n^2$? Justifique su respuesta.

c. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. ¿Cuál es la distribución de la suma $Y = \sum_{i=1}^{n} [(X_i - \mu)/\sigma]^2$? Justifique su respuesta.

90. **a.** Demuestre que $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

b. Sean X_1 y X_2 calificaciones cuantitativas y orales en un examen de aptitud, y sean Y_1 y Y_2 calificaciones correspondientes en otro examen. Si $\text{Cov}(X_1, Y_1) = 5$, $\text{Cov}(X_1, Y_2) = 1$, $\text{Cov}(X_2, Y_1) = 2$ y $\text{Cov}(X_2, Y_2) = 8$, ¿cuál es la covarianza entre las dos calificaciones totales $X_1 + X_2$ y $Y_1 + Y_2$?

91. Se selecciona al azar y se pesa dos veces un espécimen de roca de un área particular. Sea W el peso real y X_1 y X_2 los dos pesos medidos. Entonces $X_1 = W + E_1$ y $X_2 = W + E_2$, donde E_1 y E_2 son los dos errores de medición. Suponga que los E_i son independientes entre sí y de W, y que $V(E_1) = V(E_2) = \sigma_E^2$.

a. Exprese ρ, el coeficiente de correlación entre los dos pesos medidos X_1 y X_2 en función de σ_W^2, la varianza del peso real y σ_X^2, la varianza del peso medido.

b. Calcule ρ cuando $\sigma_W = 1$ kg y $\sigma_E = .01$ kg.

92. Sea A el porcentaje de un constituyente en un espécimen de roca seleccionado al azar y sea B el porcentaje de un segundo constituyente en ese mismo espécimen. Suponga que D y E son errores de medición al determinar los valores de A y B de modo que los valores medidos sean $X = A + D$ y $Y = B + E$, respectivamente. Suponga que los errores de medición son independientes entre sí y de los valores reales.

a. Demuestre que

$$\text{Corr}(X, Y) = \text{Corr}(A, B) \cdot \sqrt{\text{Corr}(X_1, X_2)} \cdot \sqrt{\text{Corr}(Y_1, Y_2)}$$

donde X_1 y X_2 son mediciones replicadas del valor de A y Y_1 y Y_2 se definen análogamente con respecto a B. ¿Qué efecto tiene la presencia del error de medición en la correlación?

b. ¿Cuál es valor máximo de $\text{Corr}(X, Y)$ cuando $\text{Corr}(X_1, X_2) = .8100$ y $\text{Corr}(Y_1, Y_2) = .9025$? ¿Es esto perturbador?

93. Sean $X_1, \ldots, X_n$ variables aleatorias independientes con valores medios $\mu_1, \ldots, \mu_n$ y varianzas $\sigma_1^2, \ldots, \sigma_n^2$. Considere una función $h(x_1, \ldots, x_n)$ y úsela para definir una nueva variable aleatoria $Y = h(X_1, \ldots, X_n)$. En condiciones un tanto generales en cuanto a la función h, si las σ_i son pequeñas con respecto a las μ_i correspondientes se puede demostrar que $E(Y) \approx h(\mu_1, \ldots, \mu_n)$ y

$$V(Y) \approx \left(\frac{\partial h}{\partial x_1}\right)^2 \cdot \sigma_1^2 + \cdots + \left(\frac{\partial h}{\partial x_n}\right)^2 \cdot \sigma_n^2$$

donde cada derivada parcial se evalúa en $(x_1, \ldots, x_n) = (\mu_1, \ldots, \mu_n)$. Suponga tres resistores con resistencias X_1, X_2, X_3 conectadas en paralelo a través de una batería con voltaje X_4. Luego, según la ley de Ohm, la corriente es

$$Y = X_4\left[\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3}\right]$$

Sean $\mu_1 = 10$ ohms, $\sigma_1 = 1.0$ ohm, $\mu_2 = 15$ ohms, $\sigma_2 = 1.0$ ohm, $\mu_3 = 20$ ohms, $\sigma_3 = 1.5$ ohms, $\mu_4 = 120$ V, $\sigma_4 = 4.0$ V. Calcule el valor esperado aproximado y la desviación estándar de la corriente (sugerido por "Random Samplings", *CHEMTECH*, 1984: 696–697).

94. Una aproximación más precisa a $E[h(X_1, \ldots, X_n)]$ en el ejercicio 93 es

$$h(\mu_1, \ldots, \mu_n) + \frac{1}{2}\sigma_1^2\left(\frac{\partial^2 h}{\partial x_1^2}\right) + \cdots + \frac{1}{2}\sigma_n^2\left(\frac{\partial^2 h}{\partial x_n^2}\right)$$

Calcule esto con $Y = h(X_1, X_2, X_3, X_4)$ dada en el ejercicio 93 y compárela con el primer término $h(\mu_1, \ldots, \mu_n)$.

95. Sean X y Y variables aleatorias normales estándar independientes y defina una nueva variable aleatoria como $U = .6X + .8Y$.

a. Determine $\text{Corr}(X, U)$.

b. ¿Cómo modificaría U para obtener $\text{Corr}(X, U) = \rho$ con un valor especificado de ρ?

96. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias que denotan n ofertas independientes para un artículo que está a la venta. Supongamos que cada X_i se distribuye uniformemente en el intervalo [100, 200]. Si el vendedor vende al mejor postor, ¿cuánto puede esperar ganar en la venta? [*Sugerencia*: sean $Y = \text{máx}(X_1, X_2, \ldots, X_n)$. En primer lugar encuentre $F_Y(y)$ al notar que $Y \leq y$ si y sólo si cada X_i es $\leq y$. Luego obtenga la función de densidad de probabilidad y $E(Y)$.]

Bibliografía

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Thomson-Brooks/Cole, Belmont, CA, 2007. Una exposición un poco más complicada de temas de probabilidad que en el presente libro.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Contiene una cuidadosa y amplia exposición de distribuciones conjuntas, reglas de probabilidad y teoremas de límites.

6 Estimación puntual

INTRODUCCIÓN

Dado un parámetro de interés, tal como la media μ o la proporción p de una población, el objetivo de la estimación puntual es utilizar una muestra para calcular un número que representa en cierto sentido una buena suposición del valor verdadero del parámetro. El número resultante se llama *estimación puntual*. En la sección 6.1 se presentan algunos conceptos generales de estimación puntual. En la sección 6.2 se describen e ilustran dos métodos importantes para obtener estimaciones puntuales: el método de momentos y el método de máxima probabilidad.

6.1 Algunos conceptos generales de estimación puntual

El objetivo de la inferencia estadística casi siempre es sacar algún tipo de conclusión sobre uno o más parámetros (características de la población). Para hacer eso un investigador tiene que obtener datos muestrales de cada una de las poblaciones estudiadas. Las conclusiones pueden entonces basarse en los valores calculados de varias cantidades muestrales. Por ejemplo, sea μ (un parámetro) la resistencia a la ruptura promedio verdadera de conexiones alámbricas utilizadas en la unión de obleas semiconductoras. Se podría tomar una muestra aleatoria de $n = 10$ conexiones y determinar la resistencia a la ruptura de cada una y se tendrían las resistencias observadas $x_1, x_2, \ldots, x_{10}$. La resistencia a la ruptura media muestral $\bar{x}$ se utilizaría entonces para sacar una conclusión con respecto al valor de μ. Asimismo, si σ^2 es la varianza de la distribución de la resistencia a la ruptura (varianza de la población, otro parámetro), el valor de la varianza muestral s^2 se utiliza para inferir algo sobre σ^2.

Cuando se discuten los métodos y conceptos generales de inferencia, es conveniente disponer de un símbolo genérico para el parámetro de interés. Se utilizará la letra griega θ para este propósito. El objetivo de la estimación puntual es seleccionar un solo número, con base en los datos muestrales, que represente un valor sensible de θ. Supóngase, por ejemplo, que el parámetro de interés es μ, la vida útil promedio verdadera de baterías de un tipo. Una muestra aleatoria de $n = 3$ baterías podría dar las vidas útiles (horas) observadas $x_1 = 5.0$, $x_2 = 6.4$, $x_3 = 5.9$. El valor calculado de la media muestral de la vida útil es $\bar{x} = 5.77$ y es razonable considerar 5.77 como un valor muy factible de μ, la "mejor suposición" del valor de μ basado en la información muestral disponible.

Supóngase que se desea estimar un parámetro de una sola población (p. ej., μ o σ) con una muestra aleatoria de tamaño n. Recuérdese por el capítulo previo de que antes de que los datos estén disponibles, las observaciones muestrales deben ser consideradas como variables aleatorias $X_1, X_2, \ldots, X_n$. Se deduce que cualquier función de las X_i, es decir, cualquier estadístico, tal como la media muestral $\bar{X}$ o la desviación estándar muestral S también es una variable aleatoria. Lo mismo es cierto si los datos disponibles se componen de más de una muestra. Por ejemplo, se pueden representar las resistencias a la tensión de m especímenes de tipo 1 y de n especímenes de tipo 2 por $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$, respectivamente. La diferencia entre las dos medias muestrales de las resistencias es $\bar{X} - \bar{Y}$, el estadístico natural para inferir sobre $\mu_1 - \mu_2$, la diferencia entre las resistencias medias de la población.

DEFINICIÓN

> Una **estimación puntual** de un parámetro θ es un número único que puede ser considerado como un valor sensible de θ. Se obtiene una estimación puntual seleccionando un estadístico apropiado y calculando su valor con los datos muestrales dados. El estadístico seleccionado se llama **estimador puntual** de θ.

En el ejemplo de las baterías que se acaba de dar, el estimador utilizado para obtener la estimación puntual de μ fue $\bar{X}$, y la estimación puntual de μ fue 5.77. Si las tres vidas útiles hubieran sido $x_1 = 5.6$, $x_2 = 4.5$ y $x_3 = 6.1$, el uso del estimador $\bar{X}$ habría dado por resultado la estimación $\bar{x} = (5.6 + 4.5 + 6.1)/3 = 5.40$. El símbolo $\hat{\theta}$ ("teta testada") se utiliza comúnmente para denotar tanto la estimación de θ como la estimación puntual que resulta de una muestra dada.* Por tanto, $\hat{\mu} = \bar{X}$ se lee como "el estimador puntual de

* Siguiendo la primera notación, se podría utilizar $\hat{\Theta}$ (una teta mayúscula) para el estimador, pero ésta es difícil de escribir.

μ es la media muestral $\overline{X}$". El enunciado "la estimación puntual de μ es 5.77" se escribe concisamente como $\hat{\mu} = 5.77$. Obsérvese que cuando se escribe $\hat{\theta} = 72.5$, no hay ninguna indicación de cómo se obtuvo esta estimación puntual (qué estadístico se utilizó). Se recomienda reportar tanto el estimador como la estimación resultante.

Ejemplo 6.1 Un fabricante automotriz ha producido un nuevo tipo de parachoques, el que se presume absorbe impactos con menos daño que los parachoques previos. El fabricante lo ha utilizado en una secuencia de 25 choques controlados contra un muro, cada uno a 10 mph, utilizando uno de sus modelos de automóvil compacto. Sea X = el número de choques que no provocaron daños visibles al automóvil. El parámetro que tiene que ser estimado es p = la proporción de todos los choques que no provocaron daños [alternativamente, $p = P$ (ningún daño en un choque)]. Si se observa que X es $x = 15$, el estimador y estimación más razonables son

$$\text{estimador } \hat{p} = \frac{X}{n} \qquad \text{estimación} = \frac{x}{n} = \frac{15}{25} = .60$$

■

Si por cada parámetro de interés hubiera sólo un estimador puntual razonable, no habría mucho para la estimación puntual. En la mayoría de los problemas, sin embargo, habrá más de un estimador razonable.

Ejemplo 6.2 Reconsidere las 20 observaciones adjuntas de voltaje de ruptura dieléctrica de piezas de resina epóxica introducidas por primera vez en el ejemplo 4.30 (sección 4.6).

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

El patrón en la gráfica de probabilidad normal dado allí es bastante recto, así que ahora se supone que la distribución de voltaje de ruptura es normal con valor medio μ. Como las distribuciones normales son simétricas, μ también es la vida útil mediana de la distribución. Se supone entonces que las observaciones dadas son el resultado de una muestra aleatoria $X_1, X_2, \ldots, X_{20}$ de esta distribución normal. Considere los siguientes estimadores y las estimaciones resultantes de μ:

a. Estimador $= \overline{X}$, estimación $= \overline{x} = \Sigma x_i/n = 555.86/20 = 27.793$

b. Estimador $= \widetilde{X}$, estimación $= \widetilde{x} = (27.94 + 27.98)/2 = 27.960$

c. Estimador $= [\text{mín}(X_i) + \text{máx}(X_i)]/2$ = el promedio de las dos vidas útiles extremas, estimación $= [\text{mín}(x_i) + \text{máx}(x_i)]/2 = (24.46 + 30.88)/2 = 27.670$

d. Estimador $= \overline{X}_{\text{rec}(10)}$, la media recortada 10% (desechar el 10% más pequeño y más grande de la muestra y luego promediar),

$$\begin{aligned}\text{estimación} &= \overline{x}_{\text{tr}(10)} \\ &= \frac{555.86 - 24.46 - 25.61 - 29.50 - 30.88}{16} \\ &= 27.838\end{aligned}$$

Cada uno de los estimadores (a)–(d) utiliza una medición diferente del centro de la muestra para estimar μ. ¿Cuál de las estimaciones se acerca más al valor verdadero? No se puede responder esta pregunta sin conocer el valor verdadero. Una pregunta que se puede hacer es: "¿cuál estimador, cuando se utiliza en otras muestras de X_i, tiende a producir estimaciones cercanas al valor verdadero? En breve se considerará este tipo de pregunta. ■

Ejemplo 6.3 El artículo "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures?" citado antes en el ejemplo 4.26, informó de las siguientes observaciones sobre X = *vacíos llenos de asfalto* (%) de 52 muestras de un cierto tipo de mezcla caliente de asfalto:

74.33	71.07	73.82	77.42	79.35	82.27	77.75	78.65	77.19
74.69	77.25	74.84	60.90	60.75	74.09	65.36	67.84	69.97
68.83	75.09	62.54	67.47	72.00	66.51	68.21	64.46	64.34
64.93	67.33	66.08	67.31	74.87	69.40	70.83	81.73	82.50
79.87	81.96	79.51	84.12	80.61	79.89	79.70	78.74	77.28
79.97	75.09	74.38	77.67	83.73	80.39	76.90		

Estimemos la varianza σ^2 de la distribución de la población. Un estimador natural es la varianza de la muestra:

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \overline{X})^2}{n - 1}$$

Minitab dio el siguiente resultado a una petición para mostrar los estadísticos descriptivos:

Variable	Count	Mean	SE Mean	StDev	Variance	Q1	Median	Q3
VFA(B)	52	73.880	0.889	6.413	41.126	67.933	74.855	79.470

Por tanto la estimación puntual de la varianza de la población es

$$\hat{\sigma}^2 = s^2 = \frac{\sum(x_i - \bar{x})^2}{52 - 1} = 41.126$$

[de forma alternativa, la fórmula de cálculo para el numerador de s^2 da

$S_{xx} = \sum x_i^2 - (\sum x_i)^2/n = 285{,}929.5964 - (3841.78)^2/52 = 2097.4124$].

Una estimación puntual de la desviación estándar de la población es entonces $\hat{\sigma} = s = \sqrt{41.126} = 6.413$.

Un estimador alternativo resulta de usar el divisor n en lugar de $n - 1$:

$$\hat{\sigma}^2 = \frac{\sum(X_i - \overline{X})^2}{n}, \quad \text{estimado} = \frac{2097.4124}{52} = 40.335$$

En breve se indicará por qué muchos estadísticos prefieren S^2 en vez de este último estimador.

El citado artículo considera ajustar cuatro distribuciones diferentes a los datos: normal, lognormal, de dos parámetros de Weibull y de tres parámetros de Weibull. Diferentes técnicas se utilizaron para concluir que los dos parámetros de Weibull proporcionan el mejor ajuste (el gráfico de probabilidad normal de los datos muestra alguna desviación de un patrón lineal). De la sección 4.5, la varianza de una variable aleatoria de Weibull es

$$\sigma^2 = \beta^2\{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2\}$$

donde α y β son los parámetros de forma y escala de la distribución. Los autores del artículo utilizaron el método de máxima verosimilitud (ver sección 6.2) para estimar estos parámetros. Las estimaciones resultantes son $\hat{\alpha} = 11.9731$, $\hat{\beta} = 77.0153$. Una estimación razonable de la varianza de la población ahora se puede obtener al sustituir las estimaciones de los dos parámetros en la expresión para σ^2; el resultado es $\hat{\sigma}^2 = 56.035$. Esta última estimación es obviamente muy diferente de la varianza de la muestra. Su validez depende de que la distribución de la población sea Weibull, mientras que la varianza de la muestra es una manera sensata para estimar σ^2 cuando hay incertidumbre en cuanto a la forma específica de la distribución de la población. ■

En el mejor de todos los mundos posibles, se podría hallar un estimador $\hat{\theta}$ con el cual $\hat{\theta} = \theta$ siempre. Sin embargo, $\hat{\theta}$ es una función de las X_i muestrales, así que es una variable aleatoria. Con algunas muestras, $\hat{\theta}$ dará un valor más grande que θ, mientras que con otras muestras $\hat{\theta}$ subestimará θ. Si se escribe

$$\hat{\theta} = \theta + \text{error de estimación}$$

entonces un estimador preciso sería uno que produzca errores de estimación pequeños, de manera que los valores estimados estén cerca del valor verdadero.

Una forma sensible de cuantificar la idea de $\hat{\theta}$ cercano a θ es considerar el error al cuadrado $(\hat{\theta} - \theta)^2$. Con algunas muestras, $\hat{\theta}$ se acercará bastante a θ y el error al cuadrado resultante se aproximará a 0. Otras muestras pueden dar valores de $\hat{\theta}$ alejados de θ, correspondientes a errores al cuadrado muy grandes. Una medida general de precisión es *la esperanza o error cuadrático medio* ECM $= E[(\hat{\theta} - \theta)^2]$. Si un primer estimador tiene una ECM más pequeña que un segundo, es natural decir que el primer estimador es el mejor. Sin embargo, el ECM en general dependerá del valor de θ. Lo que a menudo sucede es que un estimador tendrá un ECM más pequeño con algunos valores de θ y un ECM más grande con otros valores. En general no es posible determinar un estimador con el ECM más pequeño.

Una forma de librarse de este dilema es limitar la atención sólo en estimadores que tengan una propiedad deseable específica y luego determinar el mejor estimador en este grupo limitado. Una propiedad popular de esta clase en la comunidad estadística es la *ausencia de sesgo*.

Estimadores insesgados

Supóngase que se tienen dos instrumentos de medición: uno ha sido calibrado con precisión, pero el otro sistemáticamente da lecturas más pequeñas que el valor verdadero que se está midiendo. Cuando cada uno de los instrumentos se utiliza repetidamente en el mismo objeto, debido al error de medición, las mediciones observadas no serán idénticas. Sin embargo, las mediciones producidas por el primer instrumento se distribuirán en torno al valor verdadero de tal modo que en promedio este instrumento mide lo que se propone medir, por lo que se le conoce como instrumento insesgado. El segundo instrumento proporciona observaciones que tienen un componente de error o sesgo sistemático.

DEFINICIÓN

> Se dice que un estimador puntual $\hat{\theta}$ es un **estimador insesgado** de θ si $E(\hat{\theta}) = \theta$ para todo valor posible de θ. Si $\hat{\theta}$ es insesgado, la diferencia $E(\hat{\theta}) - \theta$ se conoce como el **sesgo** de $\hat{\theta}$.

Es decir, $\hat{\theta}$ es sesgado si su distribución de probabilidad (es decir, muestreo) siempre está "centrada" en el valor verdadero del parámetro. Supóngase que $\hat{\theta}$ es un estimador insesgado; entonces si $\theta = 100$, la distribución muestral $\hat{\theta}$ está centrada en 100; si $\theta = 27.5$, en ese caso la distribución muestral $\hat{\theta}$ está centrada en 27.5, y así sucesivamente. La figura 6.1 ilustra la distribución de varios estimadores sesgados e insesgados. Obsérvese que "centrada" en este caso significa que el valor esperado, no la mediana, de la distribución de $\hat{\theta}$ es igual a θ.

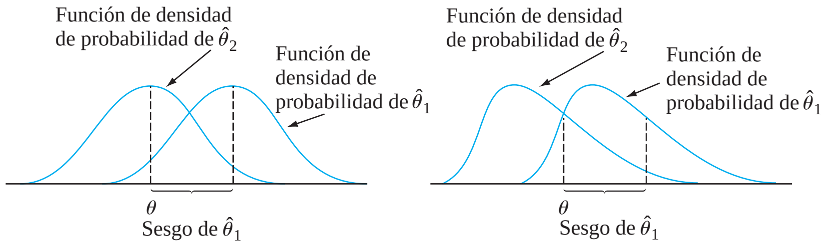


Figura 6.1 Funciones de densidad de probabilidad de un estimador sesgado $\hat{\theta}_1$ y un estimador insesgado $\hat{\theta}_2$, de un parámetro θ

Parece como si fuera necesario conocer el valor de θ (en cuyo caso la estimación es innecesaria) para ver si $\hat{\theta}$ es insesgado. Éste no suele ser el caso, sin embargo, como la ausencia de sesgo es una propiedad general de la distribución muestral del estimador, donde está centrada, lo que por lo general no depende de algún valor de parámetro particular.

En el ejemplo 6.1 se utilizó la proporción muestral X/n como estimador de p, donde X, el número de éxitos muestrales, tenía una distribución binomial con parámetros n y p. Por lo tanto

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}\,E(X) = \frac{1}{n}(np) = p$$

PROPOSICIÓN

> Cuando X es una variable aleatoria binomial con parámetros n y p, la proporción muestral $\hat{p} = X/n$ es un estimador sesgado de p.

No importa cuál sea el valor verdadero de p, la distribución del estimador $\hat{p}$ estará centrada en el valor verdadero.

Ejemplo 6.4 Suponga que X, el tiempo de reacción a un estímulo, tiene una distribución uniforme en el intervalo desde 0 hasta un límite superior desconocido θ (por tanto la función de densidad de X es de forma rectangular con altura $1/\theta$ en el intervalo $0 \le x \le \theta$). Se desea estimar θ con base en una muestra aleatoria $X_1, X_2, \ldots, X_n$ de los tiempos de reacción. Como θ es el tiempo más grande posible en toda la población de tiempos de reacción, considere como un primer estimador el tiempo de reacción muestral más grande $\hat{\theta}_1 = \text{máx}\,(X_1, \ldots, X_n)$. Si $n = 5$ y $x_1 = 4.2$, $x_2 = 1.7$, $x_3 = 2.4$, $x_4 = 3.9$, y $x_5 = 1.3$, la estimación puntual de θ es $\hat{\theta}_1 = \text{máx}(4.2, 1.7, 2.4, 3.9, 1.3) = 4.2$.

La ausencia de sesgo implica que algunas muestras darán estimaciones que exceden θ y otras que darán estimaciones más pequeñas que θ, de lo contrario θ posiblemente no podría ser el centro (punto de equilibrio) de la distribución de $\hat{\theta}_1$. Sin embargo, el estimador propuesto nunca sobrestimará θ (el valor muestral más grande no puede exceder el valor de la población más grande) y subestimará θ a menos que el valor muestral más grande sea igual a θ. Este argumento intuitivo demuestra que $\hat{\theta}_1$ es un estimador insesgado. Más precisamente, se puede demostrar (véase el ejercicio 32) que

$$E(\hat{\theta}_1) = \frac{n}{n+1}\cdot\theta < \theta \qquad \left(\text{como } \frac{n}{n+1} < 1\right)$$

El sesgo de $\hat{\theta}_1$ está dado por $n\theta/(n+1) - \theta = -\theta/(n+1)$, el cual tiende a cero a medida que n se hace grande.

Es fácil modificar $\hat{\theta}_1$ para obtener un estimador insesgado de θ. Considere el estimador

$$\hat{\theta}_2 = \frac{n+1}{n}\cdot \text{máx}\,(X_1, \ldots, X_n)$$

Utilizando este estimador en los datos se obtiene la estimación $(6/5)(4.2) = 5.04$. El hecho de que $(n+1)/n > 1$ implica que $\hat{\theta}_2$ sobrestimará θ para algunas muestras y la subestimará en otras. El valor medio de este estimador es

$$E(\hat{\theta}_2) = E\left[\frac{n+1}{n}\,\text{máx}(X_1, \ldots, X_n)\right] = \frac{n+1}{n} \cdot E\,[\text{máx}(X_1, \ldots, X_n)]$$
$$= \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta$$

Si $\hat{\theta}_2$ se utiliza repetidamente en diferentes muestras para estimar θ, algunas estimaciones serán demasiado grandes y otras demasiado pequeñas, pero a la larga no habrá ninguna tendencia simétrica a subestimar o sobrestimar θ. ■

Principio de estimación insesgada

Cuando se elige entre varios estimadores diferentes de θ, se elige uno insesgado.

De acuerdo con este principio, el estimador insesgado $\hat{\theta}_2$ en el ejemplo 6.4 deberá ser preferido al estimador sesgado $\hat{\theta}_1$. Considérese ahora el problema de estimar σ^2.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces el estimador

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \overline{X})^2}{n-1}$$

es un estimador insesgado de σ^2.

Demostración Para cualquier variable aleatoria Y, $V(Y) = E(Y^2) - [E(Y)]^2$, por lo tanto $E(Y^2) = V(Y) + [E(Y)]^2$. Aplicando esto a

$$S^2 = \frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]$$

se obtiene

$$E(S^2) = \frac{1}{n-1}\left\{\sum E(X_i^2) - \frac{1}{n}E[(\sum X_i)^2]\right\}$$
$$= \frac{1}{n-1}\left\{\sum(\sigma^2 + \mu^2) - \frac{1}{n}\{V(\sum X_i) + [E(\sum X_i)]^2\}\right\}$$
$$= \frac{1}{n-1}\left\{n\sigma^2 + n\mu^2 - \frac{1}{n}n\sigma^2 - \frac{1}{n}(n\mu)^2\right\}$$
$$= \frac{1}{n-1}\{n\sigma^2 - \sigma^2\} = \sigma^2 \quad \text{(como se desea)}$$ ■

El estimador que utiliza el divisor n se expresa como $(n-1)S^2/n$, por lo tanto

$$E\left[\frac{(n-1)S^2}{n}\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$$

Este estimador es por consiguiente sesgado. El sesgo es $(n-1)\sigma^2/n - \sigma^2 = -\sigma^2/n$. Como el sesgo es negativo, el estimador con divisor n tiende a subestimar σ^2 y por eso muchos estadísticos prefieren el divisor $n - 1$ (aunque cuando n es grande, el sesgo es pequeño y hay poca diferencia entre los dos).

Lamentablemente, el hecho de que S^2 sea insesgado para la estimación de σ^2 no implica que S sea insesgado para la estimación de σ. Sacar la raíz cuadrada estropea la propiedad de ausencia de sesgo (el valor esperado de la raíz cuadrada no es la raíz cuadrada del valor esperado). Afortunadamente, el sesgo de S es pequeño a menos que n sea muy pequeño. Hay otras buenas razones para utilizar S como estimador, especialmente cuando la distribución de la población es normal. Esto se volverá más aparente cuando se discutan los intervalos de confianza y la prueba de hipótesis en los siguientes capítulos.

En el ejemplo 6.2 se propusieron varios estimadores diferentes de la media μ de una distribución normal. Si hubiera un estimador insesgado único para μ, el problema de estimación se resolvería utilizando dicho estimador. Desafortunadamente, este no es el caso.

PROPOSICIÓN

> Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria tomada de una distribución con media μ, entonces $\overline{X}$ es un estimador sesgado de μ. Si además la distribución es continua y simétrica, entonces $\widetilde{X}$ y cualquier media recortada también son estimadores insesgados de μ.

El hecho de que $\overline{X}$ sea insesgado es simplemente un replanteamiento de una de las reglas de valor esperado: $E(\overline{X}) = \mu$ para cada valor posible de μ (para distribuciones discretas y continuas). La ausencia de sesgo de los demás estimadores es más difícil de verificar.

El siguiente ejemplo introduce otra situación en la cual existen varios estimadores insesgados para un parámetro particular.

Ejemplo 6.5 En ciertas circunstancias contaminantes orgánicos se adhieren con facilidad a las superficies de obleas y deterioran los dispositivos de fabricación de semiconductores. El artículo "Ceramic Chemical Filter for Removal of Organic Contaminants" (*J. of the Institute of Environmental Sciences and Technology*, 2003: 59–65) discutió una alternativa recientemente desarrollada de filtros de carbón convencionales para eliminar contaminación molecular transportada por el aire en aplicaciones de cuartos limpios. Un aspecto de la investigación del desempeño de filtros implicó estudiar cómo se relaciona la concentración de contaminantes en el aire con la concentración en la superficie de una oblea después de una exposición prolongada. Considere los siguientes datos representativos de x = concentración de DBP en aire y y = concentración de DBP en la superficie de obleas luego de 4 horas de exposición (ambas en μg/m^3, donde DBP = ftalato de dibutilo).

Obs. i:	1	2	3	4	5	6
x:	.8	1.3	1.5	3.0	11.6	26.6
y:	.6	1.1	4.5	3.5	14.4	29.1

Los autores comentan que la "adhesión de DBP en la superficie de obleas fue aproximadamente proporcional a la concentración de DBP en el aire". La figura 6.2 muestra una gráfica de y contra x, es decir, de los pares (x, y).

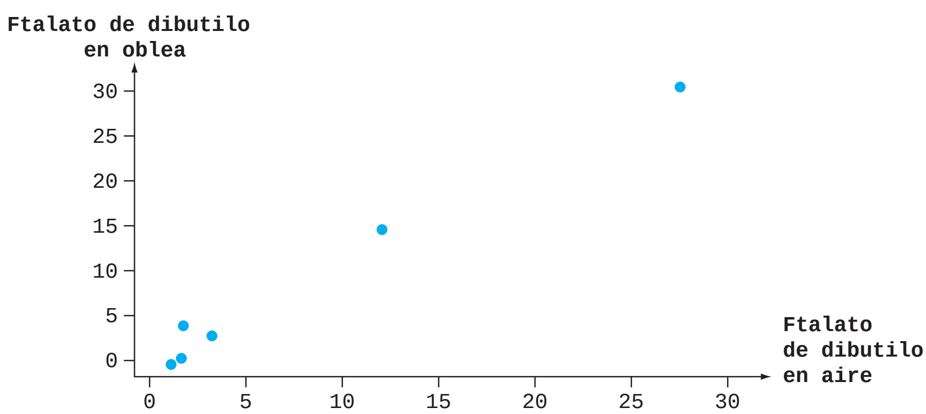


Figura 6.2 Gráfica de los datos de ftalato de dibutilo del ejemplo 6.5

Si y fuera exactamente proporcional a x, se tendría $y = \beta x$ para algún valor β, lo cual expresa que los puntos (x, y) en la gráfica quedarían exactamente sobre una línea recta con pendiente β que pasa por (0, 0). Pero esto es sólo aproximadamente el caso. Así que a continuación se supone que para cualquier x fija, la concentración de DBP en las obleas es una variable aleatoria Y con valor medio βx. Es decir, se postula que el valor *medio* de Y está relacionado con x por una línea que pasa por (0, 0) pero que el valor observado de Y en general se desviará de esta línea (esto se conoce en la literatura estadística como "regresión a través del origen").

Ahora se desea estimar el parámetro de la pendiente β. Considere los siguientes tres estimadores:

$$\#1\text{: } \hat{\beta} = \frac{1}{n}\sum \frac{Y_i}{x_i} \qquad \#2\text{: } \hat{\beta} = \frac{\sum Y_i}{\sum x_i} \qquad \#3\text{: } \hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

Las estimaciones resultantes basadas en los datos dados son 1.3497, 1.1875 y 1.1222, respectivamente. Así que de manera definitiva la estimación depende de qué estimador se utilice. Si uno de estos tres estimadores fuera insesgado y los otros dos sesgados, habría un buen motivo para utilizar el insesgado. Pero los tres son insesgados; el argumento se apoya en el hecho de que cada uno es una función lineal de las Y_i (aquí se supone que las x_i son fijas, no aleatorias):

$$E\left(\frac{1}{n}\sum \frac{Y_i}{x_i}\right) = \frac{1}{n}\sum \frac{E(Y_i)}{x_i} = \frac{1}{n}\sum \frac{\beta x_i}{x_i} = \frac{1}{n}\sum \beta = \frac{n\beta}{n} = \beta$$

$$E\left(\frac{\sum Y_i}{\sum x_i}\right) = \frac{1}{\sum x_i} E\left(\sum Y_i\right) = \frac{1}{\sum x_i}\left(\sum \beta x_i\right) = \frac{1}{\sum x_i}\beta\left(\sum x_i\right) = \beta$$

$$E\left(\frac{\sum x_i Y_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2} E\left(\sum x_i Y_i\right) = \frac{1}{\sum x_i^2}\left(\sum x_i \beta x_i\right) = \frac{1}{\sum x_i^2}\beta\left(\sum x_i^2\right) = \beta \qquad \blacksquare$$

Tanto en el ejemplo anterior como en la situación que implica estimar una media de población normal, el principio de ausencia de sesgo (preferir un estimador insesgado a uno sesgado) no puede ser invocado para seleccionar un estimador. Lo que ahora se requiere es un criterio para elegir entre estimadores sesgados.

Estimadores con varianza mínima

Supóngase que $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores de θ insesgados. Entonces, aunque la distribución de cada estimador esté centrada en el valor verdadero de θ, las dispersiones de las distribuciones en torno al valor verdadero pueden ser diferentes.

Principio de estimación insesgada con varianza mínima

Entre todos los estimadores de θ insesgados, se selecciona el de varianza mínima. El $\hat{\theta}$ resultante se llama **estimador insesgado con varianza mínima (EIVM)** de θ.

La figura 6.3 ilustra las funciones de densidad de probabilidad de los dos estimadores insesgados, donde $\hat{\theta}_1$ tiene una varianza más pequeña que $\hat{\theta}_2$. Entonces es más probable que $\hat{\theta}_1$ produzca una estimación próxima al valor verdadero θ que $\hat{\theta}_2$. El estimador insesgado con varianza mínima es, en cierto sentido, el que tiene más probabilidades entre todos los estimadores insesgados de producir una estimación cercana al verdadero θ.

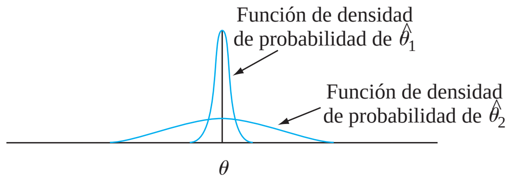


Figura 6.3 Gráficas de las funciones de densidad de probabilidad de dos estimadores insesgados diferentes

En el ejemplo 6.5, supóngase que cada Y_i está normalmente distribuida con media βx_i y varianza σ^2 (la suposición de varianza constante). Entonces se puede demostrar que el tercer estimador $\hat{\beta} = \Sigma x_i Y_i / \Sigma x_i^2$ no sólo tiene una varianza más pequeña que cualquiera de los otros dos estimadores insesgados, sino que de hecho es el estimador insesgado con varianza mínima; tiene una varianza más pequeña que *cualquier* otro estimador insesgado de β.

Ejemplo 6.6 En el ejemplo 6.4 se argumentó que cuando $X_1, \ldots, X_n$ es una muestra aleatoria tomada de una distribución uniforme en el intervalo $[0, \theta]$, el estimador

$$\hat{\theta}_1 = \frac{n+1}{n} \cdot \text{máx}\,(X_1, \ldots, X_n)$$

es insesgado para θ (previamente este estimador se denotó por $\hat{\theta}_2$). Éste no es el único estimador insesgado de θ . El valor esperado de una variable aleatoria uniformemente distribuida es simplemente el punto medio del intervalo de densidad positiva, por lo tanto $E(X_i) = \theta/2$. Esto implica que $E(\overline{X}) = \theta/2$, a partir de la cual $E(2\overline{X}) = \theta$. Es decir, el estimador $\hat{\theta}_2 = 2\overline{X}$ es insesgado para θ.

Si X está uniformemente distribuida en el intervalo de A a B, entonces $V(X) = \sigma^2 = (B - A)^2/12$. Así pues, en esta situación, $V(X_i) = \theta^2/12$, $V(\overline{X}) = \sigma^2/n = \theta^2/(12n)$ y $V(\hat{\theta}_2) = V(2\overline{X}) = 4V(\overline{X}) = \theta^2/(3n)$. Se pueden utilizar los resultados del ejercicio 32 para demostrar que $V(\hat{\theta}_1) = \theta^2/[n(n+2)]$. El estimador $\hat{\theta}_1$ tiene una varianza más pequeña que $\hat{\theta}_2$ si $3n < n(n+2)$; es decir, si $0 < n^2 - n = n(n-1)$. En tanto $n > 1$, $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, así que $\hat{\theta}_1$ es mejor estimador que $\hat{\theta}_2$. Se pueden utilizar métodos más avanzados para demostrar que $\hat{\theta}_1$ es el estimador insesgado con varianza mínima de θ; cualquier otro estimador insesgado de θ tiene una varianza que excede a $\theta^2/[n(n+2)]$. ■

Uno de los triunfos de la estadística matemática ha sido el desarrollo de una metodología para identificar el estimador insesgado con varianza mínima en una amplia variedad de situaciones. El resultado más importante de este tipo para nuestros propósitos tiene que ver con la estimación de la media μ de una distribución normal.

TEOREMA

> Sea $X_1, \ldots, X_n$ una muestra aleatoria tomada de una distribución normal con parámetros μ y σ. Entonces el estimador $\hat{\mu} = \overline{X}$ es el estimador insesgado con varianza mínima para μ.

Siempre que exista la seguridad de que la población que se está muestreando es normal, el teorema dice que $\bar{x}$ debe usarse para estimar μ. Entonces, en el ejemplo 6.2 la estimación sería $\bar{x} = 27.793$.

En algunas situaciones es posible obtener un estimador con sesgo pequeño que se preferiría al mejor estimador insesgado. Esto se ilustra en la figura 6.4. Sin embargo, los estimadores insesgados con varianza mínima a menudo son más fáciles de obtener que el tipo de estimador sesgado cuya distribución se ilustra.

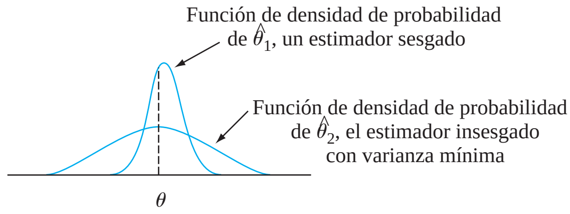


Figura 6.4 Un estimador sesgado que es preferible al estimador insesgado con varianza mínima

Algunas complicaciones

El último teorema no dice que al estimar la media μ de una población, se deberá utilizar el estimador $\overline{X}$ independientemente de la distribución que se está muestreando.

Ejemplo 6.7 Suponga que se desea estimar la conductividad térmica μ de un material. Con técnicas de medición estándar, se obtendrá una muestra aleatoria $X_1, \ldots, X_n$ de n mediciones de conductividad térmica. Suponga que la distribución de la población es un miembro de una de las siguientes tres familias:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{6.1}$$

$$f(x) = \frac{1}{\pi[1 + (x - \mu)^2]} \qquad -\infty < x < \infty \tag{6.2}$$

$$f(x) = \begin{cases} \dfrac{1}{2c} & -c \le x - \mu \le c \\ 0 & \text{de lo contrario} \end{cases} \tag{6.3}$$

La función de densidad de probabilidad (6.1) es la distribución normal, (6.2) se llama distribución de Cauchy y (6.3) es una distribución uniforme. Las tres distribuciones son simétricas con respecto a μ y de hecho la distribución de Cauchy tiene forma de campana pero con colas mucho más gruesas (más probabilidad hacia fuera) que la curva normal. La distribución uniforme no tiene colas. Los cuatro estimadores de μ considerados con anterioridad son $\overline{X}$, $\widetilde{X}$, $\overline{X}_e$ (el promedio de las dos observaciones extremas) y $\overline{X}_{\text{rec}(10)}$, una media recortada.

La muy importante moraleja en este caso es que el mejor estimador de μ depende crucialmente de qué distribución está siendo muestreada. En particular,

1. Si la muestra aleatoria proviene de una distribución normal, en ese caso $\overline{X}$ es el mejor de los cuatro estimadores, puesto que tiene una varianza mínima entre todos los estimadores insesgados.
2. Si la muestra aleatoria proviene de una distribución de Cauchy, entonces $\overline{X}$ y $\overline{X}_e$ son estimadores terribles de μ, en tanto que $\widetilde{X}$ es bastante bueno (el estimador insesgado con varianza mínima no es conocido); $\overline{X}$ es malo porque es muy sensible a las observaciones subyacentes y las colas gruesas de la distribución de Cauchy hacen que sea improbable que aparezcan tales observaciones en cualquier muestra.
3. Si la distribución subyacente es uniforme, el mejor estimador es $\overline{X}_e$; este estimador está influido en gran medida por las observaciones subyacentes, pero la carencia de colas hace que tales observaciones sean imposibles.
4. *En ninguna de estas tres situaciones es mejor la media recortada pero funciona razonablemente bien en las tres*. Es decir, $\overline{X}_{\text{rec}(10)}$ no sufre demasiado en comparación con el mejor procedimiento en cualquiera de las tres situaciones. ■

Más generalmente, investigaciones recientes en estadística han establecido que cuando se estima un punto de simetría μ de una distribución de probabilidad continua, una media recortada con proporción de recorte de 10% o 20% (para cada extremo de la muestra) produce estimaciones razonablemente comportadas dentro de un rango muy amplio de posibles modelos. Por esta razón, se dice que una media recortada con poco porcentaje de recorte es un **estimador robusto**.

En algunas situaciones, la selección no es entre dos estimadores diferentes construidos con la misma muestra, sino entre estimadores basados en dos experimentos distintos.

Ejemplo 6.8 Suponga que cierto tipo de componente tiene una distribución de vida útil exponencial con parámetro λ de modo que la vida útil esperada es $\mu = 1/\lambda$. Se selecciona una muestra de n de esos componentes y cada uno es puesto en operación. Si el experimento continúa hasta que todas las n vidas útiles, $X_1, \ldots, X_n$ han sido observadas, en ese caso $\overline{X}$ es un estimador insesgado de μ.

En algunos experimentos, sin embargo, los componentes se dejan en operación sólo hasta el tiempo de la falla r-ésima, donde $r < n$. Este procedimiento se conoce como **censura.** Sea Y_1 el tiempo de la primera falla (la vida útil mínima entre los n componentes) y Y_2 el tiempo en el cual ocurre la segunda falla (la segunda vida útil más pequeña), y así sucesivamente. Como el experimento termina en el tiempo Y_r, la vida útil acumulada al final es

$$T_r = \sum_{i=1}^{r} Y_i + (n - r)Y_r$$

A continuación se demuestra que $\hat{\mu} = T_r/r$ es un estimador insesgado de μ. Para hacerlo, se requieren dos propiedades de las variables exponenciales:

1. La propiedad de no memoria (véase la sección 4.4), la cual dice que en cualquier punto de tiempo, la vida útil restante tiene la misma distribución exponencial que la vida útil original.
2. Si $X_1, \ldots, X_k$ son independientes, cada mín$(X_1, \ldots, X_k)$ exponencialmente distribuida con parámetro λ, es exponencial con parámetro $k\lambda$ y su valor esperado es $1/(k\lambda)$.

Como los n componentes duran hasta Y_1, $n - 1$ duran una cantidad de tiempo $Y_2 - Y_1$ adicional, $n - 2$, duran una cantidad de tiempo $Y_3 - Y_2$ adicional, y así sucesivamente, otra expresión para T_r es

$$\begin{aligned} T_r = {} & nY_1 + (n-1)(Y_2 - Y_1) + (n-2)(Y_3 - Y_2) + \cdots \\ & + (n - r + 1)(Y_r - Y_{r-1}) \end{aligned}$$

Pero Y_1 es la mínima de n variables exponenciales, por tanto $E(Y_1) = 1/(n\lambda)$. Asimismo, $Y_2 - Y_1$ es la más pequeña de las $n - 1$ vidas útiles restantes, cada una exponencial con parámetro λ (por la propiedad de no memoria), así que $E(Y_2 - Y_1) = 1/[(n-1)\lambda]$. Continuando, $E(Y_{i+1} - Y_i) = 1/[(n-i)\lambda]$ así que

$$\begin{aligned} E(T_r) &= nE(Y_1) + (n-1)E(Y_2 - Y_1) + \cdots + (n - r + 1)E(Y_r - Y_{r-1}) \\ &= n \cdot \frac{1}{n\lambda} + (n-1) \cdot \frac{1}{(n-1)\lambda} + \cdots + (n-r+1) \cdot \frac{1}{(n-r+1)\lambda} \\ &= \frac{r}{\lambda} \end{aligned}$$

Por consiguiente, $E(T_r/r) = (1/r)E(T_r) = (1/r) \cdot (r/\lambda) = 1/\lambda = \mu$ como se proponía.

Como un ejemplo, supóngase que se prueban 20 componentes y $r = 10$. Entonces si los primeros diez tiempos de falla son 11, 15, 29, 33, 35, 40, 47, 55, 58 y 72, la estimación de μ es

$$\hat{\mu} = \frac{11 + 15 + \cdots + 72 + (10)(72)}{10} = 111.5$$

La ventaja del experimento con censura es que termina más rápido que el experimento sin censura. Sin embargo, se puede demostrar que $V(T_r/r) = 1/(\lambda^2 r)$, la cual es más grande que $1/(\lambda^2 n)$, la varianza de $\overline{X}$ en el experimento sin censura. ■

Reporte de una estimación puntual: el error estándar

Además de reportar el valor de una estimación puntual, se debe dar alguna indicación de su precisión. La medición usual de precisión es el error estándar del estimador usado.

DEFINICIÓN

El **error estándar** de un estimador $\hat{\theta}$ es su desviación estándar $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. Éste es la magnitud de una desviación típica o representativa entre una estimación y el valor de θ. Si el error estándar implica parámetros desconocidos cuyos valores pueden ser estimados, la sustitución de estas estimaciones en $\sigma_{\hat{\theta}}$ da el **error estándar estimado** (desviación estándar estimada) del estimador. El error estándar estimado puede ser denotado por $\hat{\sigma}_{\hat{\theta}}$ (el ^ sobre σ recalca que $\sigma_{\hat{\theta}}$ está siendo estimada) o por $s_{\hat{\theta}}$.

Ejemplo 6.9 (Continuación del ejemplo 6.2)

Suponiendo que el voltaje de ruptura está normalmente distribuido, $\hat{\mu} = \overline{X}$ es la mejor estimación de μ. Si se sabe que el valor de σ es 1.5, el error estándar de $\overline{X}$ es $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1.5/\sqrt{20} = .335$. Si, como casi siempre es el caso, el valor de σ es desconocido, la estimación $\hat{\sigma} = s = 1.462$ se sustituye en $\sigma_{\overline{X}}$ para obtener el error estándar estimado $\hat{\sigma}_{\overline{X}} = s_{\overline{X}} = s/\sqrt{n} = 1.462/\sqrt{20} = .327$. ■

Ejemplo 6.10 (Continuación del ejemplo 6.1)

El error estándar de $\hat{p} = X/n$ es

$$\sigma_{\hat{p}} = \sqrt{V(X/n)} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

Como p y $q = 1 - p$ son desconocidas (o bien ¿por qué estimarlas?), se sustituyen $\hat{p} = x/n$ y $\hat{q} = 1 - x/n$ en $\sigma_{\hat{p}}$, para obtener el error estándar estimado $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{(.6)(.4)/25} = .098$. Alternativamente, como el valor más grande de pq se obtiene cuando $p = q = .5$, un límite superior en el error estándar es $\sqrt{1/(4n)} = .10$. ■

Cuando la distribución del estimador puntual $\hat{\theta}$ es normal de modo aproximado, lo que a menudo será el caso cuando n es grande, entonces se puede estar confiado de manera razonable en que el valor verdadero de θ queda dentro de aproximadamente 2 errores estándar (desviaciones estándar) de $\hat{\theta}$. De este modo, si una muestra de $n = 36$ vidas útiles de componentes da $\hat{\mu} = \overline{x} = 28.50$ y $s = 3.60$, por consiguiente $s/\sqrt{n} = .60$, dentro de dos errores estándar estimados, $\hat{\mu}$ se traslada al intervalo $28.50 \pm (2)(.60) = (27.30, 29.70)$.

Si $\hat{\theta}$ no es necesariamente normal en forma aproximada pero es insesgado, entonces se puede demostrar que la estimación se desviará de θ hasta por 4 errores estándar cuando mucho 6% del tiempo. Se esperaría entonces que el valor verdadero quede dentro de 4 errores estándar de $\hat{\theta}$ (y ésta es una afirmación muy conservadora, puesto que se aplica a *cualquier* $\hat{\theta}$ insesgado). Resumiendo, el error estándar indica de forma aproximada a qué distancia de $\hat{\theta}$ se puede esperar que quede el valor verdadero de θ.

La forma del estimador de $\hat{\theta}$ puede ser suficientemente complicada de modo que la teoría estadística estándar no pueda ser aplicada para obtener una expresión para $\sigma_{\hat{\theta}}$. Esto es cierto, por ejemplo, en el caso $\theta = \sigma$, $\hat{\theta} = S$; la desviación estándar del estadístico S, σ_S, en general no puede ser determinada. No hace mucho se introdujo un método de computadora intensivo llamado *bootstrap* para abordar este problema. Supóngase que la función de densidad de probabilidad de la población es $f(x; \theta)$, un miembro de una familia paramétrica particular, y que los datos $x_1, x_2, \ldots, x_n$ dan $\hat{\theta} = 21.7$. Ahora se utiliza la computadora para obtener "muestras bootstrap" tomadas de la función de densidad de probabilidad $f(x; 21.7)$, y por cada muestra se calcula una "estimación bootstrap" $\hat{\theta}^*$:

Primera muestra "bootstrap" : $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_1^*$
Segunda muestra "bootstrap": $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_2^*$
$\vdots$
B-ésima muestra bootstrap: $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_B^*$

A menudo se utiliza $B = 100$ o 200. Ahora sea $\bar{\theta}^* = \Sigma\hat{\theta}_i^*/B$, la media muestral de las estimaciones "bootstrap". La **estimación bootstrap** del error estándar de $\hat{\theta}$ ahora es simplemente la desviación estándar muestral de los $\hat{\theta}_i^*$:

$$S_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\Sigma(\hat{\theta}_i^* - \bar{\theta}^*)^2}$$

(En la literatura de bootstrap, a menudo se utiliza B en lugar de $B - 1$; con valores típicos de B, casi siempre hay poca diferencia entre las estimaciones resultantes.)

Ejemplo 6.11 Un modelo teórico sugiere que X, el tiempo para la ruptura de un fluido aislante entre electrodos a un voltaje particular, tiene $f(x; \lambda) = \lambda e^{-\lambda x}$, una distribución exponencial. Una muestra aleatoria de $n = 10$ tiempos de ruptura (minutos) da los datos siguientes:

41.53 18.73 2.99 30.34 12.33 117.52 73.02 223.63 4.00 26.78

Como $E(X) = 1/\lambda$, $E(\bar{X}) = 1/\lambda$, una estimación razonable de λ es $\hat{\lambda} = 1/\bar{x} = 1/55.087 = .018153$. Se utilizaría entonces un programa de computadora estadístico para obtener $B = 100$ muestras bootstrap, cada una de tamaño 10, provenientes de $f(x; .018153)$. La primera muestra fue 41.00, 109.70, 16.78, 6.31, 6.76, 5.62, 60.96, 78.81, 192.25, 27.61, con la cual $\sum x_i^* = 545.8$ y $\hat{\lambda}_1^* = 1/54.58 = .01832$. El promedio de las 100 estimaciones bootstrap es $\bar{\lambda}^* = .02153$, y la desviación estándar muestral de estas 100 estimaciones es $s_{\hat{\lambda}} = .0091$, la estimación bootstrap del error estándar de $\hat{\lambda}$. Un histograma de los 100$\hat{\lambda}_i$* resultó un tanto positivamente asimétrico, lo que sugiere que la distribución muestral de $\hat{\lambda}$ también tiene esta propiedad. ■

En ocasiones un investigador desea estimar una característica poblacional sin suponer que la distribución de la población pertenece a una familia paramétrica particular. Una instancia de esto ocurrió en el ejemplo 6.7, cuando una media recortada 10% fue propuesta para estimar el centro θ de una distribución de población simétrica. Los datos del ejemplo 6.2 dieron $\hat{\theta} = \bar{x}_{\text{rec}(10)} = 27.838$, pero ahora no hay ninguna $f(x; \theta)$ supuesta, por consiguiente, ¿cómo se puede obtener una muestra bootstrap? La respuesta es considerar que la muestra constituye la población (las $n = 20$ observaciones en el ejemplo 6.2) y tome B muestras diferentes, cada una de tamaño n, *con* reemplazo, de esta población. El libro de Bradley Efron y Robert Tibshirani o el de John Rice incluidos en la bibliografía del capítulo proporcionan más información.

EJERCICIOS Sección 6.1 (1–19)

1. Los datos adjuntos sobre resistencia a la flexión (MPa) de vigas de concreto de un tipo se introdujeron en el ejemplo 1.2.

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

a. Calcule una estimación puntual del valor medio de resistencia de la población conceptual de todas las vigas fabricadas de esta manera, y diga qué estimador utilizó. [*Sugerencia*: $\Sigma x_i = 219.8$.]

b. Calcule una estimación puntual del valor de resistencia que separa el 50% más débil de dichas vigas del 50% más resistente, y diga qué estimador utilizó.

c. Calcule e interprete una estimación puntual de la desviación estándar de la población σ. ¿Qué estimador utilizó? [*Sugerencia*: $\sum x_i^2 = 1860.94$.]

d. Calcule una estimación puntual de la proporción de las vigas cuya resistencia a la flexión exceda de 10 MPa. [*Sugerencia*: considere una observación como "éxito" si excede de 10.]

e. Calcule una estimación puntual del coeficiente de variación de la población σ/μ y diga qué estimador utilizó.

2. Una muestra de 20 estudiantes que recientemente tomaron un curso de estadística elemental arrojó la siguiente información sobre la marca de calculadora que poseían. (T = Texas Instruments, H = Hewlett Packard, C = Casio, S = Sharp):

T	T	H	T	C	T	T	S	C	H
S	S	T	H	C	T	T	T	H	T

a. Estime la verdadera proporción de los estudiantes que poseen una calculadora Texas Instruments.

b. De los 10 estudiantes que poseían una calculadora TI, 4 tenían calculadoras graficadoras. Estime la proporción de estudiantes que no poseen una calculadora graficadora TI.

3. Considere la siguiente muestra de observaciones sobre el espesor de recubrimiento de pintura de baja viscosidad ("Achieving a Target Value for a Manufacturing Process: A Case Study", *J. of Quality Technology*, 1992: 22–26):

.83	.88	.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

Suponga que la distribución del espesor de recubrimiento es normal (una gráfica de probabilidad normal soporta fuertemente esta suposición).

a. Calcule la estimación puntual del valor medio del espesor de recubrimiento y diga qué estimador utilizó.

b. Calcule una estimación puntual de la mediana de la distribución del espesor de recubrimiento y diga qué estimador utilizó.

c. Calcule la estimación puntual del valor que separa el 10% más grande de todos los valores de la distribución del espesor del 90% restante y diga qué estimador utilizó. [*Sugerencia*: exprese lo que está tratando de estimar en función de μ y σ.]

d. Estime $P(X < 1.5)$; es decir, la proporción de todos los valores de espesor menores que 1.5. [*Sugerencia*: si conociera los valores de μ y σ podría calcular esta probabilidad. Estos valores no están disponibles, pero pueden ser estimados.]

e. ¿Cuál es el error estándar estimado del estimador que utilizó en el inciso (b)?

4. El artículo del cual se extrajeron los datos en el ejercicio 1 también dio las observaciones de resistencias adjuntas de cilindros:

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

Antes de obtener los datos, denote las resistencias de vigas mediante $X_1, \ldots, X_m$ y las resistencias de cilindros $Y_1, \ldots, Y_n$. Suponga que las X_i constituyen una muestra aleatoria de una distribución con media μ_1 y desviación estándar σ_1 y que las Y_i forman una muestra aleatoria (independiente de las X_i) de otra distribución con media μ_2 y desviación estándar σ_2.

a. Use las reglas de valor esperado para demostrar que $\overline{X} - \overline{Y}$ es un estimador insesgado de $\mu_1 - \mu_2$. Calcule el estimador para los datos dados.

b. Use las reglas de varianza del capítulo 5 para obtener una expresión para la varianza y desviación estándar (error estándar) del estimador del inciso (a) y luego calcule el error estándar estimado.

c. Calcule una estimación puntual de la razón σ_1/σ_2 de las dos desviaciones estándar.

d. Suponga que se seleccionan al azar una sola viga y un solo cilindro. Calcule una estimación puntual de la varianza de la diferencia $X - Y$ entre la resistencia de las vigas y la resistencia de los cilindros.

5. Como ejemplo de una situación en la que varios estadísticos diferentes podrían ser razonablemente utilizados para calcular una estimación puntual, considere una población de N facturas. Asociado con cada factura se encuentra su "valor en libros", la cantidad anotada de dicha factura. Sea T el valor en libros total, una cantidad conocida. Algunos de estos valores en libros son erróneos. Se realizará una auditoría seleccionando al azar n facturas y determinando el valor auditado (correcto) para cada una. Suponga que la muestra da los siguientes resultados (en dólares).

	Factura				
	1	**2**	**3**	**4**	**5**
Valor en libros	300	720	526	200	127
Valor auditado	300	520	526	200	157
Error	0	200	0	0	−30

Sea

$\overline{Y}$ = valor en libros medio muestral

$\overline{X}$ = valor auditado medio muestral

$\overline{D}$ = error medio muestral

Proponga tres estadísticos diferentes para estimar el valor total (correcto) auditado: uno que implique exactamente N y $\overline{X}$, otro que implique T, N y $\overline{D}$ y el último que implique T y $\overline{X}/\overline{Y}$. Si $N = 5000$ y $T = 1{,}761{,}300$, calcule las tres estimaciones puntuales correspondientes. (El artículo "Statistical Models and Analysis in Auditing", *Statistical Science*, 1989: 2-33, discute las propiedades de estos tres estimadores.)

6. Considere las observaciones adjuntas sobre los flujos de una corriente de agua (miles de acres-pies) registradas en una estación en Colorado durante el periodo del 1 de abril al 31 de agosto durante 31 años (tomadas de un artículo que apareció en el volumen 1974 de *Water Resources Research*).

127.96	210.07	203.24	108.91	178.21
285.37	100.85	89.59	185.36	126.94
200.19	66.24	247.11	299.87	109.64
125.86	114.79	109.11	330.33	85.54
117.64	302.74	280.55	145.11	95.36
204.91	311.13	150.58	262.09	477.08
94.33				

Una gráfica de probabilidad apropiada soporta el uso de la distribución lognormal (véase la sección 4.5) como modelo razonable del flujo de corriente de agua.

a. Calcule los parámetros de la distribución [*Sugerencia*: recuerde que X tiene una distribución lognormal con parámetros μ y σ^2 si $\ln(X)$ está normalmente distribuida con media μ y varianza σ^2.]

b. Use las estimaciones del inciso (a) para calcular una estimación del valor esperado del flujo de corriente de agua [*Sugerencia*: ¿Cuál es $E(X)$?]

7. **a.** Se selecciona una muestra aleatoria de 10 casas en un área particular, cada una de las cuales se calienta con gas natural y se determina la cantidad de gas (terms) utilizada por cada casa durante el mes de enero. Las observaciones resultantes son 103, 156, 118, 89, 125, 147, 122, 109, 138, 99. Sea μ el consumo de gas promedio durante enero de todas las casas del área. Calcule una estimación puntual de μ.

b. Suponga que hay 10,000 casas en esta área que utilizan gas natural para calefacción. Sea τ la cantidad total de gas consumido por todas estas casas durante enero. Calcule τ con los datos del inciso (a). ¿Qué estimador utilizó para calcular su estimación?

c. Use los datos del inciso (a) para estimar p, la proporción de todas las casas que usaron por lo menos 100 terms.

d. Proporcione una estimación puntual de la mediana de la población usada (el valor intermedio en la población de todas las casas) basada en la muestra del inciso (a). ¿Qué estimador utilizó?

8. En una muestra aleatoria de 80 componentes de un tipo, se encontraron 12 defectuosos.

a. Dé una estimación puntual de la proporción de todos los componentes que *no* están defectuosos.

b. Se tiene que construir un sistema seleccionando al azar dos de estos componentes y conectándolos en serie, como se muestra a continuación.

La conexión en serie implica que el sistema funcionará si y sólo si ningún componente esté defectuoso (es decir, ambos componentes funcionan apropiadamente). Estime la proporción de todos los sistemas que funcionan de manera apropiada. [*Sugerencia*: si p denota la probabilidad de que el componente funcione apropiadamente, ¿cómo puede ser expresada P(el sistema funciona) en función de p?]

9. Se examina cada uno de 150 artículos recién fabricados y se anota el número de rayones por artículo (se supone que los artículos están libres de rayones) y se obtienen los siguientes datos:

Número de rayones por artículo	0	1	2	3	4	5	6	7
Frecuencia observada	18	37	42	30	13	7	2	1

Sea X = el número de rayones en un artículo seleccionado al azar y suponga que X tiene una distribución de Poisson con parámetro μ.

a. Determine un estimador insesgado de μ y calcule la estimación de los datos. [*Sugerencia*: $E(X) = \mu$ para una distribución de Poisson de X, por lo tanto $E(\overline{X}) = ?$]

b. ¿Cuál es la desviación estándar (error estándar) de su estimador? Calcule el error estándar estimado. [*Sugerencia*: $\sigma_X^2 = \mu$ con distribución de Poisson de X.]

10. Con una larga varilla de longitud μ se va a trazar una gráfica cuadrada en la cual la longitud de cada lado es μ. Por consiguiente el área de la curva será μ^2. Sin embargo, no se conoce el valor de μ así que decide hacer n mediciones independientes $X_1, X_2, \ldots, X_n$ de la longitud. Suponga que cada X_i tiene una media μ (mediciones insesgadas) y varianza σ^2.

a. Demuestre que $\overline{X}^2$ no es un estimador insesgado de μ^2. [*Sugerencia*: con cualquier variable aleatoria Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Aplique ésta con $Y = \overline{X}$.]

b. ¿Para qué valor de el estimador $\overline{X}^2 - kS^2$ k es insesgado para μ^2? [*Sugerencia*: calcule $E(\overline{X}^2 - kS^2)$.]

11. De n_1 varones fumadores seleccionados al azar, X_1 fuman cigarrillos con filtro, mientras que de n_2 fumadoras seleccionadas al azar, X_2 fuman cigarrillos con filtro. Sean p_1 y p_2 las probabilidades de que un varón y una mujer seleccionados al azar fumen, respectivamente, cigarrillos con filtro.

a. Demuestre que $(X_1/n_1) - (X_2/n_2)$ es un estimador insesgado de $p_1 - p_2$. [*Sugerencia*: $E(X_i) = n_i p_i$ con $i = 1, 2$.]

b. ¿Cuál es el error estándar del estimador en el inciso (a)?

c. ¿Cómo utilizaría los valores observados x_1 y x_2 para estimar el error estándar de su estimador?

d. Si $n_1 = n_2 = 200$, $x_1 = 127$ y $x_2 = 176$, use el estimador del inciso (a) para obtener una estimación de $p_1 - p_2$.

e. Use el resultado del inciso (c) y los datos del inciso (d) para estimar el error estándar del estimador.

12. Suponga que un tipo de fertilizante rinde μ_1 por acre con varianza σ^2, mientras que el rendimiento esperado de un segundo tipo de fertilizante es μ_2, con la misma varianza σ^2. Sean S_1^2 y S_2^2 las varianzas muestrales de los rendimientos basadas en tamaños muestrales n_1 y n_2, respectivamente, de los dos fertilizantes. Demuestre que el estimador combinado

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

es un estimador insesgado de σ^2.

13. Considere una muestra aleatoria $X_1, \ldots, X_n$ de la función de densidad de probabilidad

$$f(x;\theta) = .5(1 + \theta x) \qquad -1 \le x \le 1$$

donde $-1 \le \theta \le 1$ (esta distribución se presenta en la física de partículas). Demuestre que $\hat{\theta} = 3\overline{X}$ es un estimador insesgado de θ. [*Sugerencia*: primero determine $\mu = E(X) = E(\overline{X})$.]

14. Una muestra de n aviones de combate Pandemonium capturados tienen los números de serie $x_1, x_2, x_3, \ldots, x_n$. La CIA sabe que los aviones fueron numerados consecutivamente en la fábrica comenzando con α y terminando con β, por lo que el número total de aviones fabricados es $\beta - \alpha + 1$ (p. ej., si $\alpha = 17$ y $\beta = 29$, entonces $29 - 17 + 1 = 13$ aviones con números de serie 17, 18, 19, . . . , 28, 29 fueron fabricados). Sin embargo, la CIA no conoce los valores de α o β. Un estadístico de la CIA sugiere utilizar el estimador $\text{máx}(X_i) - \text{mín}(X_i) + 1$ para estimar el número total de aviones fabricados.

a. Si $n = 5$, $x_1 = 237$, $x_2 = 375$, $x_3 = 202$, $x_4 = 525$ y $x_5 = 418$, ¿cuál es la estimación correspondiente?

b. ¿En qué condiciones de la muestra será el valor de la estimación exactamente igual al número total verdadero de aviones? ¿Alguna vez será más grande la estimación que el total verdadero? ¿Piensa que el estimador es insesgado para estimar $\beta - \alpha + 1$? Explique en una o dos oraciones.

15. Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria tomada de una distribución de Rayleigh con función de densidad de probabilidad

$$f(x; \theta) = \frac{x}{\theta} e^{-x^2/(2\theta)} \qquad x > 0$$

a. Se puede demostrar que $E(X^2) = 2\theta$. Use este hecho para construir un estimador insesgado de θ basado en $\sum X_i^2$ (y use reglas de valor esperado para demostrar que es insesgado).

b. Calcule θ a partir de las siguientes $n = 10$ observaciones de esfuerzo vibratorio de un aspa de turbina en condiciones específicas:

16.88	10.23	4.59	6.66	13.68
14.23	19.87	9.40	6.51	10.95

16. Suponga que el crecimiento promedio verdadero μ de un tipo de planta durante un periodo de 1 año es idéntico al de un segundo tipo, aunque la varianza del crecimiento del primer tipo es σ^2, en tanto que para el segundo tipo la varianza es $4\sigma^2$. Sean $X_1, \ldots, X_m$, m observaciones de crecimiento independientes del primer tipo [por consiguiente $E(X_i) = \mu$, $V(X_i) = \sigma^2$], y sean $Y_1, \ldots, Y_n$, n observaciones de crecimiento independientes del segundo tipo [$E(Y_i) = \mu$, $V(Y_i) = 4\sigma^2$].

a. Demuestre que para cualquier δ entre 0 y 1, el estimador $\hat{\mu} = \delta\overline{X} + (1 - \delta)\overline{Y}$ es insesgado para μ.

b. Con m y n fijas, calcule $V(\hat{\mu})$ y luego determine el valor de δ que reduzca al mínimo $V(\hat{\mu})$. [*Sugerencia*: diferencie $V(\hat{\mu})$ con respecto a δ.]

17. En el capítulo 3 se definió una variable aleatoria binomial negativa como el número de fallas que ocurren antes del r-ésimo éxito en una secuencia de ensayos con éxitos y fallos independientes e idénticos. La función de masa de probabilidad (fmp) de X es

$$nb(x; r, p) = \binom{x + r - 1}{x} p^r (1 - p)^x \qquad x = 0, 1, 2, \ldots$$

a. Suponga que $r \geq 2$. Demuestre que

$$\hat{p} = (r - 1)/(X + r - 1)$$

es un estimador insesgado de p. [*Sugerencia*: escriba $E(\hat{p})$ y elimine $x + r - 1$ dentro de la suma.]

b. Un reportero desea entrevistar a cinco individuos que apoyan a un candidato y comienza preguntándoles si (S) o no (F) apoyan al candidato. Si la secuencia de respuestas es *SFFSFFFSSS*, estime p = la proporción verdadera que apoya al candidato.

18. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad $f(x)$ que es simétrica con respecto a μ, de modo que $\widetilde{X}$ es un estimador insesgado de μ. Si n es grande, se puede demostrar que $V(\widetilde{X}) \approx 1/(4n[f(\mu)]^2)$.

a. Compara $V(\widetilde{X})$ con $V(\overline{X})$ cuando la distribución subyacente es normal.

b. Cuando la función de densidad de probabilidad subyacente es de Cauchy (véase el ejemplo 6.7), $V(\overline{X}) = \infty$, por lo tanto $\overline{X}$ es un estimador terrible. ¿Cuál es $V(\widetilde{X})$ en este caso cuando n es grande?

19. Una investigadora desea estimar la proporción de estudiantes en una universidad que han violado el código de honor. Habiendo obtenido una muestra aleatoria de n estudiantes, se da cuenta de que si a cada uno le pregunta "¿has violado el código de honor?" probablemente recibirá algunas respuestas faltas de veracidad. Considere el siguiente esquema, conocido como técnica de **respuesta aleatorizada.** La investigadora forma un mazo de 100 cartas de las cuales 50 son de tipo I y 50 de tipo II.

Tipo I: ¿Has violado el código de honor (sí o no)?

Tipo II: ¿El último dígito de su número telefónico es un 0, 1 o 2 (sí o no)?

A cada estudiante en la muestra aleatoria se le pide que baraje el mazo, que saque una carta y que responda la pregunta con sinceridad. A causa de la pregunta irrelevante en las cartas de tipo II, una respuesta afirmativa ya no estigmatiza al que responde, así que se supone que éste es sincero. Sea p la proporción de violadores del código de honor (es decir, la probabilidad de que un estudiante seleccionado al azar sea un violador) y sea $\lambda = P(\text{respuesta sí})$. Entonces λ y p están relacionados por $\lambda = .5p + (.5)(.3)$.

a. Sea Y el número de respuestas afirmativa, por consiguiente $Y \sim \text{Bin}(n, \lambda)$. Por tanto Y/n es un estimador insesgado de λ. Deduzca un estimador de p basado en Y. Si $n = 80$ y $y = 20$, ¿cuál es su estimación? [*Sugerencia*: despeje p de $\lambda = .5p + .15$ y luego sustituya Y/n en lugar de λ.]

b. Use el hecho de que $E(Y/n) = \lambda$ para demostrar que su estimador $\hat{p}$ es insesgado.

c. Si hubiera 70 cartas de tipo I y 30 de tipo II, ¿cuál sería su estimador para p?

6.2 Métodos de estimación puntual

La definición de ausencia de sesgo no indica en general cómo se pueden obtener los estimadores insesgados. A continuación se discuten dos métodos "constructivos" para obtener estimadores puntuales: el método de momentos y el método de máxima probabilidad. Por *constructivo* se quiere dar a entender que la definición general de cada tipo de estimador sugiere explícitamente cómo obtener el estimador en cualquier problema específico. Aun cuando se prefieren los estimadores de máxima probabilidad a los de momento debido a ciertas propiedades de eficiencia, a menudo requieren significativamente más cálculo que los estimadores de momento. En ocasiones es el caso que estos métodos dan estimadores insesgados.

El método de momentos

La idea básica de este método es poder igualar ciertas características muestrales, tales como la media, a los valores esperados de la población correspondiente. Luego al resolver estas ecuaciones para valores de parámetros desconocidos se obtienen los estimadores.

DEFINICIÓN

Si $X_1, \ldots, X_n$ constituyen una muestra aleatoria proveniente de una función de masa de probabilidad o de una función de densidad de probabilidad $f(x)$. Con $k = 1, 2, 3, \ldots$, el **momento *k*-ésimo de la población** o el **momento *k*-ésimo de la distribución *f*(*x*),** es $E(X^k)$. El **momento muestral *k*-ésimo** es $(1/n)\sum_{i=1}^{n} X_i^k$.

Por consiguiente, el primer momento de la población es $E(X) = \mu$ y el primer momento muestral es $\Sigma X_i/n = \overline{X}$. Los segundos momentos de la población y muestral son $E(X^2)$ y $\sum X_i^2/n$, respectivamente. Los momentos de la población serán funciones de cualesquiera parámetros desconocidos $\theta_1, \theta_2, \ldots$.

DEFINICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con función de masa de probabilidad o función de densidad de probabilidad $f(x; \theta_1, \ldots, \theta_m)$, donde $\theta_1, \ldots, \theta_m$ son parámetros cuyos valores son desconocidos. Entonces los **estimadores de momento** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ se obtienen igualando los primeros m momentos muestrales con los primeros m momentos de la población correspondientes y resolviendo para $\theta_1, \ldots, \theta_m$.

Si, por ejemplo, $m = 2$, $E(X)$ y $E(X^2)$ serán funciones de θ_1 y θ_2. Con $E(X) = (1/n)\,\Sigma X_i$ $(= \overline{X})$ y $E(X^2) = (1/n)\Sigma X_i^2$ se obtienen dos ecuaciones con θ_1 y θ_2. La solución define entonces los estimadores.

Ejemplo 6.12 Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de tiempos de servicio de n clientes en una instalación, donde la distribución subyacente se supone exponencial con el parámetro λ. Como sólo hay un parámetro que tiene que ser estimado, el estimador se obtiene igualando $E(X)$ a $\overline{X}$. Como $E(X) = 1/\lambda$ con una distribución exponencial, ésta da $1/\lambda = \overline{X}$ o $\lambda = 1/\overline{X}$. El estimador de momento de λ es entonces $\hat{\lambda} = 1/\overline{X}$. ■

Ejemplo 6.13 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución gamma con parámetros α y β. De acuerdo con la sección 4.4, $E(X) = \alpha\beta$ y $E(X^2) = \beta^2\Gamma(\alpha + 2)/\Gamma(\alpha) = \beta^2(\alpha + 1)\alpha$. Los estimadores de momento de α y β se obtienen resolviendo

$$\overline{X} = \alpha\beta \qquad \frac{1}{n}\Sigma X_i^2 = \alpha(\alpha + 1)\beta^2$$

Como $\alpha(\alpha + 1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$ y la primera ecuación implica $\alpha^2\beta^2 = \overline{X}^2$, la segunda ecuación se vuelve

$$\frac{1}{n}\Sigma X_i^2 = \overline{X}^2 + \alpha\beta^2$$

Ahora si se divide cada miembro de esta segunda ecuación entre el miembro correspondiente de la primera ecuación y se sustituye otra vez se obtienen los estimadores

$$\hat{\alpha} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{\beta} = \frac{(1/n)\sum X_i^2 - \overline{X}^2}{\overline{X}}$$

Para ilustrar, los datos de tiempo de sobrevivencia mencionados en el ejemplo 4.24 son

152	115	109	94	88	137	152	77	160	165
125	40	128	123	136	101	62	153	83	69

con $\bar{x} = 113.5$ y $(1/20)\sum x_i^2 = 14{,}087.8$. Los estimadores son

$$\hat{\alpha} = \frac{(113.5)^2}{14{,}087.8 - (113.5)^2} = 10.7 \qquad \hat{\beta} = \frac{14{,}087.8 - (113.5)^2}{113.5} = 10.6$$

Estas estimaciones de α y β difieren de los valores sugeridos por Gross y Clark porque ellos utilizaron una técnica de estimación diferente. ■

Ejemplo 6.14 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución binomial negativa generalizada con parámetros r y p (vea la sección 3.5). Como $E(X) = r(1-p)/p$ y $V(X) = r(1-p)/p^2$, $E(X^2) = V(X) + [E(X)]^2 = r(1-p)(r - rp + 1)/p^2$. Si se iguala $E(X)$ a $\overline{X}$ y $E(X^2)$ a $(1/n)\sum X_i^2$ a la larga se obtiene

$$\hat{p} = \frac{\overline{X}}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{r} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2 - \overline{X}}$$

Como ilustración, Reep, Pollard y Benjamin ("Skill and Chance in Ball Games", *J. of Royal Stat. Soc.*, 1971: 623–629) consideran la distribución binomial negativa como modelo del número de goles por juego anotados por los equipos de la Liga Nacional de Jockey. Los datos de 1966–1967 son los siguientes (420 juegos):

Goles	0	1	2	3	4	5	6	7	8	9	10
Frecuencia	29	71	82	89	65	45	24	7	4	1	3

Entonces,

$$\bar{x} = \Sigma x_i/420 = [(0)(29) + (1)(71) + \cdots + (10)(3)]/420 = 2.98$$

y

$$\sum x_i^2/420 = [(0)^2(29) + (1)^2(71) + \cdots + (10)^2(3)]/420 = 12.40$$

Por consiguiente,

$$\hat{p} = \frac{2.98}{12.40 - (2.98)^2} = .85 \qquad \hat{r} = \frac{(2.98)^2}{12.40 - (2.98)^2 - 2.98} = 16.5$$

Aunque r por definición debe ser positivo, el denominador de $\hat{r}$ podría ser negativo, lo que indica que la distribución binomial negativa no es apropiada (o que el estimador de momento es defectuoso). ■

Estimación de máxima probabilidad

El método de máxima probabilidad lo introdujo por primera vez R. A. Fisher, genetista y estadístico en la década de 1920. La mayoría de los estadísticos recomiendan este método, por lo menos cuando el tamaño de muestra es grande, puesto que los estimadores resultantes tienen ciertas propiedades de eficiencia deseables (véase la proposición en la página 262).

Ejemplo 6.15 Se obtuvo una muestra de diez cascos de ciclista nuevos fabricados por una compañía. Al probarlos, se encontró que el primero, el tercero y el décimo estaban agrietados, en tanto que los demás no. Sea $p = P$(casco agrietado) es decir, p es la proporción de todos los cascos que están agrietados. Defina variables aleatorias (de Bernoulli) $X_1, X_2, \ldots, X_{10}$ como

$$X_1 = \begin{cases} 1 \text{ si el primer casco está agrietado} \\ 0 \text{ si el primer casco no está agrietado} \end{cases} \cdots X_{10} = \begin{cases} 1 \text{ si el décimo casco está agrietado} \\ 0 \text{ si el décimo casco no está agrietado} \end{cases}$$

Entonces, para la muestra obtenida, $X_1 = X_3 = X_{10} = 1$ y las otras siete X_i son cero. La función de masa de probabilidad de cualquier X_i particular es $p^{x_i}(1 - p)^{1-x_i}$ que se convierte en p si $x_i = 1$ y $1 - p$ cuando $x_i = 0$. Supóngase ahora que las condiciones de los diferentes cascos son independientes una de la otra. Esto implica que las X_i son independientes, por lo que su función de masa de probabilidad conjunta es el producto de las funciones de masa de probabilidad individuales. Así que la función de masa de probabilidad conjunta de las X_i observadas es

$$f(x_1, \ldots, x_{10}; p) = p(1 - p)p \cdots p = p^3(1 - p)^7 \tag{6.4}$$

Supóngase que $p = .25$. Entonces la probabilidad de observar la muestra que en realidad se obtiene es $(.25)^3(.75)^7 = .002086$. Si en cambio $p = .50$, entonces esta probabilidad es $(.50)^3(.50)^7 = .000977$. ¿Para qué valor de p es más probable que la muestra observada haya ocurrido? Es decir, ¿para qué valor de p es la función de masa de probabilidad conjunta (6.4) tan grande como puede ser? ¿Qué valor de p maximiza (6.4)? La figura 6.5(a) muestra un gráfico de la *probabilidad* (6.4) en función de p. Parece que la gráfica alcanza su pico por encima de $p = .3 =$ la proporción de los cascos defectuosos en la muestra. La figura 6.5(b) muestra un gráfico del logaritmo natural de (6.4) ya que $ln[g(u)]$ es una función estrictamente creciente de $g(u)$, encontrar u para maximizar la función $g(u)$ es lo mismo que encontrar u para maximizar $ln[g(u)]$.

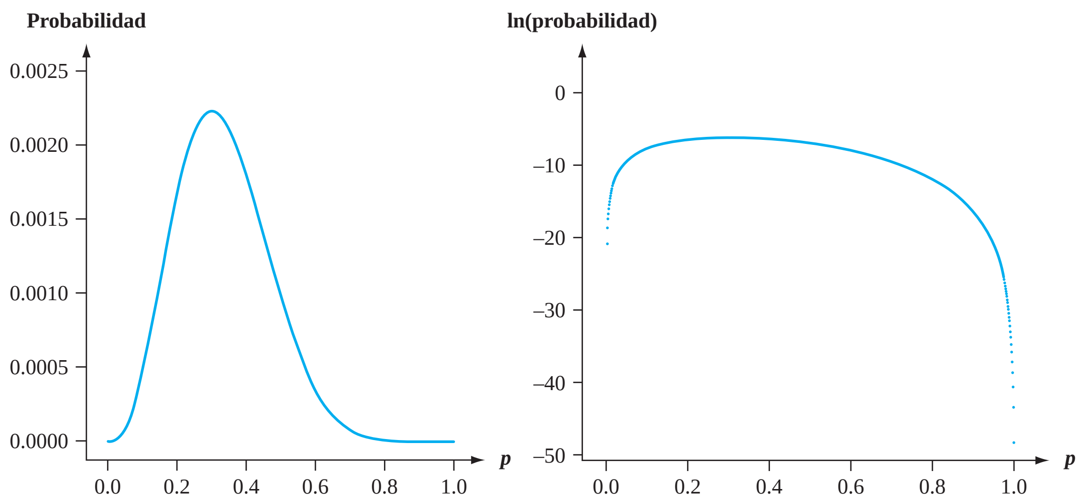


Figura 6.5 (a) Gráfica de la probabilidad (función de masa de probabilidad conjunta) (6.4) del ejemplo 6.15; (b) Gráfica del logaritmo natural de la probabilidad

Podemos comprobar nuestra impresión visual usando el cálculo para hallar el valor de p que maximiza (6.4). Trabajar con el logaritmo natural de la función de masa de probabilidad conjunta suele ser más fácil que trabajar con la función de masa de probabilidad conjunta, como esta última es típicamente un producto, su logaritmo será una suma. Aquí

$$\ln[f(x_1, \ldots, x_{10}; p)] = \ln[p^3(1 - p)^7] = 3\ln(p) + 7\ln(1 - p) \tag{6.5}$$

Por tanto

$$\frac{d}{dp}\{\ln[f(x_1, \ldots, x_{10}; p)]\} = \frac{d}{dp}\{3\ln(p) + 7\ln(1 - p)\} = \frac{3}{p} + \frac{7}{1 - p}(-1)$$
$$= \frac{3}{p} - \frac{7}{1 - p}$$

[el (-1) viene de la regla de la cadena en el cálculo]. Igualando esta derivada a 0 y despejando p da $3(1 - p) = 7p$, de lo cual $3 = 10p$ y así $p = 3/10 = .30$ como conjeturamos. Es decir, nuesta estimación puntual es $\hat{p} = .30$. Se llama *estimación de máxima verosimilitud*, ya que es el valor del parámetro que maximiza la probabilidad (fmp conjunta) de la muestra observada. En general, la segunda derivada debe ser examinada para asegurarse de que se haya obtenido un máximo, pero aquí esto es obvio en la figura 6.5.

Supongamos que en vez de decirle la condición de cada casco, sólo había sido informado de que tres de los diez eran defectuosos. Entonces tendríamos el valor observado de una variable aleatoria binomial X = el número de cascos defectuosos. La función de masa de probabilidad de X es $\binom{10}{x} p^x (1 - p)^{10-x}$. Para $x = 3$, esto se convierte en $\binom{10}{3}p^3(1 - p)^7$. El coeficiente binomial $\binom{10}{3}$ es irrelevante para la maximización, así que nuevamente $\hat{p} = .30$. ■

DEFINICIÓN

Sean $X_1, X_2, \ldots, X_n$ que tienen una función de masa de probabilidad o una función de densidad de probabilidad

$$f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m) \tag{6.6}$$

donde los parámetros $\theta_1, \ldots, \theta_m$ tienen valores desconocidos. Cuando $x_1, \ldots, x_n$ son los valores muestrales observados y (6.6) se considera como una función de $\theta_1, \ldots, \theta_m$, se llama **función de probabilidad.** Las estimaciones de máxima probabilidad $\hat{\theta}_1, \ldots, \hat{\theta}_m$ son aquellos valores de las θ_i que incrementan al máximo la función de probabilidad, de modo que

$$f(x_1, \ldots, x_n; \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \text{ para todas las } \theta_1, \ldots, \theta_m$$

Cuando se sustituyen las X_i en lugar de las x_i, se obtienen los **estimadores de máxima probabilidad**.

La función de probabilidad dice qué tan probable es que la muestra observada sea una función de los posibles valores de parámetro. Al incrementarse al máximo la probabilidad se obtienen los valores de parámetro con los que es más probable que la muestra observada haya sido generada; es decir, los valores de parámetro que "más concuerdan" con los datos observados.

Ejemplo 6.16 Suponga que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución exponencial con parámetro λ. Debido a la independencia, la función de probabilidad es un producto de las funciones de densidad de probabilidad individuales:

$$f(x_1, \ldots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \Sigma x_i}$$

El logaritmo natural de la función de probabilidad es

$$\ln[f(x_1, \ldots, x_n; \lambda)] = n \ln(\lambda) - \lambda \Sigma x_i$$

Si se iguala $(d/d\lambda)$[ln(probabilidad)] a cero se obtiene $n/\lambda - \Sigma x_i = 0$ o $\lambda = n/\Sigma x_i = 1/\bar{x}$. Por consiguiente el estimador de máxima probabilidad es $\hat{\lambda} = 1/\bar{X}$; es idéntico al método de estimador de momentos [pero no es un estimador insesgado, puesto que $E(1/\bar{X}) \neq 1/E(\bar{X})$]. ■

Ejemplo 6.17 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal. La función de probabilidad es

$$f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/(2\sigma^2)} \cdot \cdots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/(2\sigma^2)}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\Sigma(x_i-\mu)^2/(2\sigma^2)}$$

por consiguiente

$$\ln[f(x_1, \ldots, x_n; \mu, \sigma^2)] = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$$

Para determinar los valores maximizantes de μ y σ^2, se deben sacar las derivadas parciales de $\ln(f)$ con respecto a μ y σ^2, igualarlas a cero y resolver las dos ecuaciones resultantes. Omitiendo los detalles, los estimadores de máxima probabilidad resultantes son

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{\sum(X_i - \overline{X})^2}{n}$$

El estimador de máxima probabilidad de σ^2 no es el estimador insesgado, por consiguiente dos principios diferentes de estimación (ausencia de sesgo y máxima probabilidad) dan dos estimadores diferentes. ■

Ejemplo 6.18 En el capítulo 3 se mencionó el uso de la distribución de Poisson para modelar el número de "eventos" que ocurren en una región bidimensional. Suponga que cuando la región R se está muestreando tiene área $a(R)$, el número X de eventos que ocurren en R tiene una distribución de Poisson con parámetro $\lambda a(R)$ (donde λ es el número esperado de eventos por unidad de área) y que las regiones no traslapantes dan X independientes.

Suponga que un ecólogo selecciona n regiones no traslapantes $R_1, \ldots, R_n$ y cuenta el número de plantas de una especie en cada región. La función de masa de probabilidad (verosimilitud) conjunta es entonces

$$p(x_1, \ldots, x_n; \lambda) = \frac{[\lambda \cdot a(R_1)]^{x_1} e^{-\lambda \cdot a(R_1)}}{x_1!} \cdot \cdots \cdot \frac{[\lambda \cdot a(R_n)]^{x_n} e^{-\lambda \cdot a(R_n)}}{x_n!}$$
$$= \frac{[a(R_1)]^{x_1} \cdot \cdots \cdot [a(R_n)]^{x_n} \cdot \lambda^{\Sigma x_i} \cdot e^{-\lambda\Sigma a(R_i)}}{x_1! \cdot \cdots \cdot x_n!}$$

El ln(probabilidad) es

$$\ln[p(x_1, \ldots, x_n; \lambda)] = \sum x_i \cdot \ln[a(R_i)] + \ln(\lambda) \cdot \sum x_i - \lambda\sum a(R_i) - \sum \ln(x_i!)$$

Tomando $d/d\lambda \ln(p)$ e igualándola a cero da

$$\frac{\sum x_i}{\lambda} - \sum a(R_i) = 0$$

por consiguiente

$$\lambda = \frac{\sum x_i}{\sum a(R_i)}$$

El estimador de máxima probabilidad es entonces $\hat{\lambda} = \sum X_i / \sum a(R_i)$. Ésta es razonablemente intituitiva porque λ es la densidad verdadera (plantas por unidad de área), mientras que $\hat{\lambda}$ es la densidad muestral puesto que $\Sigma a(R_i)$ es tan sólo el área total muestreada. Como $E(X_i) = \lambda \cdot a(R_i)$, el estimador es insesgado.

En ocasiones se utiliza un procedimiento de muestreo alternativo. En lugar de fijar las regiones que van a ser muestreadas, el ecólogo seleccionará n puntos en toda la región de interés y sea y_i = la distancia del punto i-ésimo a la planta más cercana. La función de distribución acumulativa de Y = distancia a la planta más cercana es

$$F_Y(y) = P(Y \le y) = 1 - P(Y > y) = 1 - P\left(\begin{array}{c}\text{ninguna planta en}\\ \text{un círculo de radio } y\end{array}\right)$$

$$= 1 - \frac{e^{-\lambda\pi y^2}(\lambda\pi y^2)^0}{0!} = 1 - e^{-\lambda \cdot \pi y^2}$$

Al sacar la derivada de $F_Y(y)$ con respecto a y resulta

$$f_Y(y; \lambda) = \begin{cases} 2\pi\lambda y e^{-\lambda\pi y^2} & y \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si ahora se forma la probabilidad $f_Y(y_1; \lambda) \cdot \ldots \cdot f_Y(y_n; \lambda)$, se diferencia ln(probabilidad), y así sucesivamente, el estimador de máxima probabilidad resultante es

$$\hat{\lambda} = \frac{n}{\pi \sum Y_i^2} = \frac{\text{número de plantas observadas}}{\text{área total muestreada}}$$

la que también es una densidad muestral. Se puede demostrar que un ambiente ralo (pequeño λ), el método de distancia es mejor en cierto sentido, en tanto que en un ambiente denso, el primer método de muestreo es mejor. ■

Ejemplo 6.19 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad de Weibull

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-(x/\beta)^\alpha} & x \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si se escribe la probabilidad y el ln(probabilidad) y luego $(\partial/\partial\alpha)[\ln(f)] = 0$ y $(\partial/\partial\beta)[\ln(f)] = 0$ se obtienen las ecuaciones

$$\alpha = \left[\frac{\sum x_i^\alpha \cdot \ln(x_i)}{\sum x_i^\alpha} - \frac{\sum \ln(x_i)}{n}\right]^{-1} \qquad \beta = \left(\frac{\sum x_i^\alpha}{n}\right)^{1/\alpha}$$

Estas dos ecuaciones no pueden ser resueltas explícitamente para obtener fórmulas generales de los estimadores de máxima probabilidad $\hat{\alpha}$ y $\hat{\beta}$. En su lugar, por cada muestra $x_1, \ldots, x_n$, las ecuaciones deben ser resueltas con un procedimiento numérico iterativo. Incluso los estimadores de momento de α y β son un tanto complicados (véase el ejercicio 21). ■

Estimación de funciones de parámetros

En el ejemplo 6.17, se obtuvo el estimador de máxima probabilidad de σ^2 cuando la distribución subyacente es normal. El estimador de máxima probabilidad de $\sigma = \sqrt{\sigma^2}$ como el de muchos otros estimadores de máxima probabilidad, es fácil de obtener con la siguiente proposición.

PROPOSICIÓN

El principio de invarianza

Sean $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ los estimadores de máxima probabilidad de los parámetros $\theta_1, \theta_2, \ldots, \theta_m$. Entonces el estimador de máxima probabilidad de cualquier función $h(\theta_1, \theta_2, \ldots, \theta_m)$ de estos parámetros es la función $h(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m)$ de los estimadores de máxima probabilidad.

Ejemplo 6.20 (Continuación del ejemplo 6.17)

En el caso normal, los estimadores de máxima probabilidad de μ y σ^2 son $\hat{\mu} = \overline{X}$ y $\hat{\sigma}^2 = \sum(X_i - \overline{X})^2/n$. Para obtener el estimador de máxima probabilidad de la función $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, sustituya los estimadores de máxima probabilidad en la función:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum(X_i - \overline{X})^2\right]^{1/2}$$

El estimador de máxima probabilidad de σ no es la desviación estándar muestral S, aunque se aproximan bastante a menos que n sea bastante pequeño. ■

Ejemplo 6.21 (Continuación del ejemplo 6.19)

El valor medio de una variable aleatoria X que tiene una distribución de Weibull es

$$\mu = \beta \cdot \Gamma(1 + 1/\alpha)$$

El estimador de máxima probabilidad de μ es por consiguiente $\hat{\mu} = \hat{\beta}\Gamma(1 + 1/\hat{\alpha})$ donde $\hat{\alpha}$ y $\hat{\beta}$ son los estimadores de máxima probabilidad de α y β. En particular, $\overline{X}$ no es el estimador de máxima probabilidad de μ, aunque es un estimador insesgado. Por lo menos para n grande, $\hat{\mu}$ es un mejor estimador que $\overline{X}$.

Para los datos que figuran en el ejemplo 6.3, los estimadores de máxima probabilidad de los parámetros de Weibull son $\hat{\alpha} = 11.9731$ y $\hat{\beta} = 77.0153$, de los cuales $\hat{\mu} = 73.80$. Esta estimación está muy cerca de la media de la muestra 73.88. ■

Comportamiento del estimador de máxima probabilidad con muestra grande

Aunque el principio de la estimación de máxima probabilidad tiene un considerable atractivo intuitivo, la siguiente proposición proporciona razones adicionales fundamentales para el uso de estimadores de máxima probabilidad.

PROPOSICIÓN

En condiciones muy generales en relación con la distribución conjunta de la muestra, cuando el tamaño n de la muestra es grande, el estimador de máxima probabilidad de cualquier parámetro θ es aproximadamente insesgado $[E(\hat{\theta}) \approx \theta]$ y su varianza es casi tan pequeña como la que puede ser lograda por cualquier estimador. Expresado de otra manera, el estimador de máxima probabilidad $\hat{\theta}$ es aproximadamente el estimador insesgado con varianza mínima de θ.

Debido a este resultado y al hecho de que las técnicas basadas en el cálculo casi siempre pueden ser utilizadas para obtener los estimadores de máxima probabilidad (aunque a veces se requieren métodos numéricos, tales como el método de Newton), la estimación de máxima probabilidad es la técnica de estimación más ampliamente utilizada entre los estadísticos. Muchos de los estimadores utilizados en lo que resta del libro son estimadores de máxima probabilidad. Sin embargo, la obtención de un estimador de máxima probabilidad requiere que se especifique la distribución subyacente.

Algunas complicaciones

En ocasiones no se puede utilizar el cálculo para obtener estimadores de máxima probabilidad.

Ejemplo 6.22

Suponga que mi tiempo de espera de un autobús está uniformemente distribuido en $[0, \theta]$ y que se observaron los resultados $x_1, \ldots, x_n$ de una muestra aleatoria tomada de esta distribución. Como $f(x; \theta) = 1/\theta$ con $0 \leq x \leq \theta$ y 0 de lo contrario,

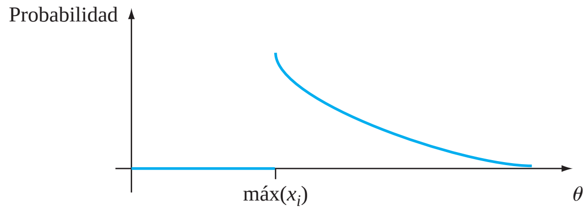


Figura 6.6 Función de probabilidad para el ejemplo 6.22

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 \le x_1 \le \theta, \ldots, 0 \le x_n \le \theta \\ 0 & \text{de lo contrario} \end{cases}$$

En tanto máx(x_i) $\le \theta$, la probabilidad es $1/\theta^n$, la cual es positiva, pero en cuanto $\theta <$ máx(x_i), la probabilidad se reduce a 0. Esto se ilustra en la figura 6.6. El cálculo no funciona porque el máximo de la probabilidad ocurre en un punto de discontinuidad, pero la figura indica que $\hat{\theta} =$ máx(X_i). Por consiguiente si mis tiempos de espera son 2.3, 3.7, 1.5, .4 y 3.2, entonces el estimador de máxima probabilidad es $\hat{\theta} = 3.7$. De acuerdo con el ejemplo 6.4, el estimador de máxima probabilidad no es insesgado. ■

Ejemplo 6.23 Un método que a menudo se utiliza para estimar el tamaño de una población de vida silvestre implica realizar un experimento de captura-recaptura. En este experimento se captura una muestra inicial de M animales, cada uno se éstos se etiqueta y luego son regresados a la población. Tras permitir un tiempo suficiente para que los individuos etiquetados se mezclen con la población, se captura otra muestra de tamaño n. Con $X =$ el número de animales etiquetados en la segunda muestra, el objetivo es utilizar las x observadas para estimar la población de tamaño N.

El parámetro de interés es $\theta = N$, el cual puede asumir sólo valores enteros, así que incluso después de determinar la función de probabilidad (función de masa de probabilidad de X en este caso), el uso del cálculo para obtener N presentaría dificultades. Si se considera un éxito la recaptura de un animal previamente etiquetado, entonces el muestreo es sin reemplazo de una población que contiene M éxitos y $N - M$ fallas, de modo que X es una variable aleatoria hipergeométrica y la función de probabilidad es

$$p(x; N) = h(x; n, M, N) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

No obstante la naturaleza de valor entero de N, sería difícil evaluar la derivada de $p(x; N)$. Sin embargo, si se considera la razón de $p(x; N)$ y $p(x; N - 1)$, se tiene

$$\frac{p(x; N)}{p(x; N-1)} = \frac{(N-M) \cdot (N-n)}{N(N-M-n+x)}$$

Esta razón es más grande que 1 si y sólo si $N < Mn/x$. El valor de N con el cual $p(x; N)$ se incrementa al máximo es por consiguiente el entero más grande menor que Mn/x. Si se utiliza la notación matemática estándar $[r]$ para el entero más grande menor o igual a r, el estimador de máxima probabilidad de N es $\hat{N} = [Mn/x]$. Como ilustración, si $M = 200$ peces se sacan de un lago y etiquetan, posteriormente $n = 100$ son recapturados y entre los 100 hay $x = 11$ etiquetados, en ese caso $\hat{N} = [(200)(100)/11] = [1818.18] = 1818$. La estimación es en realidad un tanto intuitiva; x/n es la proporción de la muestra recapturada etiquetada, mientras que M/N es la proporción de toda la población etiquetada. La estimación se obtiene igualando estas dos proporciones (estimando una proporción poblacional mediante una proporción muestral). ■

Supóngase que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una función de densidad de probabilidad $f(x; \theta)$ simétrica con respecto a θ aunque el investigador no está seguro de la forma de la función f. Es entonces deseable utilizar un estimador $\hat{\theta}$ *robusto*; es decir, uno que funcione bien con una amplia variedad de funciones de densidad de probabilidad subyacentes. Un estimador como ése es una media recortada. En años recientes, los estadísticos han propuesto otro tipo de estimador, llamado *estimador M*, basado en una generalización de la estimación de máxima probabilidad. En lugar de incrementar al máximo el logaritmo de la probabilidad $\Sigma \ln[f(x; \theta)]$ para una f específica, se incrementa al máximo $\Sigma\rho(x_i; \theta)$. Se selecciona la "función objetivo" ρ para que dé un estimador con buenas propiedades de robustez. El libro de David Hoaglin y colaboradores (véase la bibliografía) contiene una buena exposición de esta materia.

EJERCICIOS Sección 6.2 (20–30)

20. Una prueba de diagnóstico para una enfermedad determinada se aplica a n individuos de los que se sabe que no tienen la enfermedad. Sea $X =$ el número uno de los n resultados de prueba que son positivos (lo que indica la presencia de la enfermedad, por lo que X es el número de falsos positivos) y $p =$ probabilidad de que el resultado de un individuo de prueba libre de la enfermedad es positivo (es decir, p es la verdadera proporción de resultados de las pruebas de individuos libres de enfermedades que son positivos). Supongamos que sólo X está disponible en lugar de la secuencia real de los resultados de la prueba.

a. Derive el estimador de máxima probabilidad de p. Si $n = 20$ y $x = 3$, ¿cuál es la estimación?

b. ¿Es insesgado el estimador del inciso (a)?

c. Si $n = 20$ y $x = 3$, ¿cuál es el estimador de máxima probabilidad $(1 - p)^5$ que ninguna de las próximas cinco pruebas realizadas en los individuos libres de la enfermedad sean positivas?

21. Si X tiene una distribución de Weibull con parámetros α y β, entonces

$$E(X) = \beta \cdot \Gamma(1 + 1/\alpha)$$
$$V(X) = \beta^2\{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2\}$$

a. Basado en una muestra aleatoria $X_1, \ldots, X_n$, escriba ecuaciones para el método de estimadores de momentos de β y α. Demuestre que una vez que se obtiene la estimación de α, la estimación de β se puede hallar en una tabla de la función gamma, y que la estimación de α es la solución de una ecuación complicada que implica la función gamma.

b. Si $n = 20$, $\bar{x} = 28.0$ y $\sum x_i^2 = 16{,}500$, calcule la estimaciones. [*Sugerencia*: $[\Gamma(1.2)]^2/\Gamma(1.4) = .95$.]

22. Sea X la proporción de tiempo asignado que un estudiante seleccionado al azar pasa resolviendo cierta prueba de aptitud. Suponga que la función de densidad de probabilidad de X es

$$f(x; \theta) = \begin{cases} (\theta + 1)x^\theta & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

donde $-1 < \theta$. Una muestra aleatoria de diez estudiantes produce los datos $x_1 = .92$, $x_2 = .79$, $x_3 = .90$, $x_4 = .65$, $x_5 = .86$, $x_6 = .47$, $x_7 = .73$, $x_8 = .97$, $x_9 = .94$, $x_{10} = .77$.

a. Use el método de momentos para obtener un estimador de θ y luego calcule la estimación para estos datos.

b. Obtenga el estimador de máxima probabilidad de θ y luego calcule la estimación para los datos dados.

23. Dos sistemas de computadoras diferentes son supervisados durante un total de n semanas. Sea X_i el número de descomposturas del primer sistema durante la semana i-ésima y suponga que las X_i son independientes y que se extraen de una distribución de Poisson con parámetro μ_1. Asimismo, sea Y_i el número de descomposturas del segundo sistema durante la semana i-ésima y suponga independencia con cada Y_i extraída de una distribución de Poisson con parámetro μ_2. Deduzca los estimadores de máxima probabilidad de μ_1, μ_2 y $\mu_1 - \mu_2$. [*Sugerencia*: utilizando independencia, escriba la función de masa de probabilidad conjunta de las X_i y Y_i juntas.]

24. Un vehículo con un defecto particular en su sistema de control de emisiones es llevado a una serie de mecánicos seleccionados al azar hasta que $r = 3$ de ellos han diagnosticado correctamente el problema. Supongamos que esto requiere de los diagnósticos de 20 mecánicos diferentes (por lo que hubo 17 diagnósticos incorrectos). Sea $p = P$ (diagnóstico correcto), por lo que el estimador de máxima probabilidad es la proporción de todos los mecánicos que bien podría diagnosticar el problema. ¿Cuál el estimador de máxima probabilidad de p? ¿Es el mismo estimador de máxima probabilidad si una muestra aleatoria de 20 mecánicos resulta en tres diagnósticos correctos? Explique. ¿Cómo funciona el estimador de máxima probabilidad en comparación con la estimación resultante de la utilización del estimador imparcial dada en el ejercicio 17?

25. Se determina la resistencia al esfuerzo cortante de soldaduras de 10 puntos de prueba y se obtienen los siguientes datos (lb/pulg^2):

392	376	401	367	389	362	409	415	358	375

a. Suponiendo que la resistencia al esfuerzo cortante está normalmente distribuida, estime la resistencia al esfuerzo cortante promedio verdadera y la desviación estándar de la resistencia al esfuerzo cortante utilizando el método de máxima probabilidad.

b. De nuevo suponiendo una distribución normal, calcule el valor de resistencia por debajo del cual 95% estarán las resistencias de todas las soldaduras. [*Sugerencia*: ¿cuál es el 95º percentil en función de μ y σ? Utilice ahora el principio de invarianza.]

26. Remítase al ejercicio 25. Suponga que decide examinar otra soldadura de puntos de prueba. Sea $X =$ resistencia al esfuerzo cortante de la soldadura. Use los datos dados para obtener el estimador de máxima probabilidad de $P(X \leq 400)$. [*Sugerencia*: $P(X \leq 400) = \Phi((400 - \mu)/\sigma)$.]

27. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución gamma con parámetros α y β.

a. Deduzca las ecuaciones cuyas soluciones dan los estimadores de máxima probabilidad de α y β. ¿Piensa que pueden ser resueltos explícitamente?

b. Demuestre que el estimador de máxima probabilidad de $\mu = \alpha\beta$ es $\hat{\mu} = \overline{X}$.

28. Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de la distribución de Rayleigh con función de densidad dada en el ejercicio 15, determine:

a. El estimador de máxima probabilidad de θ y luego calcule la estimación para los datos de esfuerzo de vibración dados en ese ejercicio. ¿Es este estimador el mismo estimador insesgado sugerido en el ejercicio 15?

b. El estimador de máxima probabilidad de la mediana de la distribución del esfuerzo de vibración. [*Sugerencia*: exprese primero la mediana en función de θ.]

29. Considere la muestra aleatoria $X_1, X_2, \ldots, X_n$ de la función de densidad de probabilidad exponencial desplazada

$$f(x; \lambda, \theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x \geq \theta \\ 0 & \text{de lo contrario} \end{cases}$$

Con $\theta = 0$ da la función de densidad de probabilidad de la distribución exponencial considerada previamente (con densidad positiva a la derecha de cero). Un ejemplo de la distribución exponencial desplazada apareció en el ejemplo 4.5, en el cual la variable de interés fue el tiempo entre vehículos en el flujo de tránsito, y $\theta = .5$ fue el tiempo entre vehículos mínimo posible.

a. Obtenga los estimadores de máxima probabilidad de θ y λ.

b. Si $n = 10$ observaciones de tiempo entre vehículos son realizadas y se obtienen los siguientes resultados 3.11, .64, 2.55, 2.20, 5.44, 3.42, 10.39, 8.93, 17.82 y 1.30, calcule las estimaciones de θ y λ.

30. En el instante $t = 0$, 20 componentes idénticos son puestos a prueba. La distribución de la vida útil de cada uno es exponencial con parámetro λ. El experimentador deja la instalación de prueba sin supervisar. A su regreso 24 horas más tarde, el experimentador termina de inmediato la prueba después de notar que $y = 15$ de los 20 componentes aún están en operación (así que 5 han fallado). Obtenga el estimador de máxima probabilidad de λ. [*Sugerencia*: sea $Y =$ el número que sobrevive 24 horas. En ese caso $Y \sim \text{Bin}(n, p)$. ¿Cuál es el estimador de máxima probabilidad de p? Observe ahora que $p = P(X_i \geq 24)$, donde X_i está exponencialmente distribuida. Esto relaciona λ con p, de modo que el primero puede ser estimado una vez que lo ha sido el segundo.]

EJERCICIOS SUPLEMENTARIOS (31–38)

31. Se dice que un estimador $\hat{\theta}$ es **consistente** si con cualquier $\epsilon > 0$, $P(|\hat{\theta} - \theta| \geq \epsilon) \to 0$ a medida que $n \to \infty$. Es decir, $\hat{\theta}$ es consistente si, a medida que el tamaño de muestra se hace más grande, es menos y menos probable que $\hat{\theta}$ se aleje más que ϵ del valor verdadero de θ. Demuestre que $\overline{X}$ es un estimador consistente de μ cuando $\sigma^2 < \infty$ mediante la desigualdad de Chebyshev del ejercicio 44 del capítulo 3. [*Sugerencia*: la desigualdad puede ser reescrita en la forma

$$P(|Y - \mu_Y| \geq \epsilon) \leq \sigma_Y^2/\epsilon$$

Ahora identifique Y con $\overline{X}$.]

32. a. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución uniforme en $[0, \theta]$. Entonces el estimador de máxima probabilidad de θ es $\hat{\theta} = Y = \text{máx}(X_i)$. Use el hecho de que $Y \leq y$ si y sólo si cada $X_i \leq y$ para deducir la función de distribución acumulativa de Y. Luego demuestre que la función de densidad de probabilidad de $Y = \text{máx}(X_i)$ es

$$f_Y(y) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \theta \\ 0 & \text{de lo contrario} \end{cases}$$

b. Use el resultado del inciso (a) para demostrar que el estimador de máxima probabilidad es sesgado pero que $(n + 1)\,\text{máx}(X_i)/n$ es insesgado.

33. En el instante $t = 0$, hay un individuo vivo en una población. Un **proceso de nacimientos puro** se desarrolla entonces como sigue. El tiempo hasta que ocurre el primer nacimiento está exponencialmente distribuido con parámetro λ. Después del primer nacimiento, hay dos individuos vivos. El tiempo hasta que el primero da a luz otra vez es exponencial con parámetro λ y del mismo modo para el segundo individuo. Por consiguiente, el tiempo hasta el siguiente nacimiento es el mínimo de dos variables (λ) exponenciales, el cual es exponencial con parámetro 2λ. Asimismo, una vez que el segundo nacimiento ha ocurrido, hay tres individuos vivos, de modo que el tiempo hasta el siguiente nacimiento es una variable aleatoria exponencial con parámetro 3λ, y así sucesivamente (aquí se está utilizando la propiedad de no memoria de la distribución exponencial). Suponga que se observa el proceso hasta que el sexto nacimiento ha ocurrido y los tiempos hasta los nacimientos sucesivos son 25.2, 41.7, 51.2, 55.5, 59.5, 61.8 (con los cuales deberá calcular los tiempos entre nacimientos sucesivos). Obtenga el estimador de máxima probabilidad de λ. [*Sugerencia*: la probabilidad es un producto de términos exponenciales.]

34. El **error cuadrático medio** de un estimador $\hat{\theta}$ es $\text{ECM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$. Si $\hat{\theta}$ es insesgado, entonces $\text{ECM}(\hat{\theta}) = V(\hat{\theta})$, pero en general $\text{ECM}(\hat{\theta}) = V(\hat{\theta}) + (\text{sesgo})^2$. Considere el estimador $\hat{\sigma}^2 = KS^2$, donde $S^2 =$ varianza muestral. ¿Qué valor de K reduce al mínimo el error cuadrático medio de este estimador cuando la distribución de la población es normal? [*Sugerencia*: se puede demostrar que

$$E[(S^2)^2] = (n + 1)\sigma^4/(n - 1)$$

En general, es difícil determinar $\hat{\theta}$ para reducir al mínimo el ECM($\hat{\theta}$), por lo cual se buscan sólo estimadores insesgados y se reduce al mínimo $V(\hat{\theta})$.]

35. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad simétrica con respecto a μ. Un estimador de μ que se ha visto que funciona bien con una amplia variedad de distribuciones subyacentes es el *estimador de Hodges-Lehmann*. Para definirlo, primero calcule para cada $i \le j$ y cada $j = 1, 2, \ldots, n$ el promedio por pares $\overline{X}_{i,j} = (X_i + X_j)/2$. Entonces el estimador es $\hat{\mu}$ = la mediana de las $\overline{X}_{i,j}$. Calcule el valor de esta estimación con los datos del ejercicio 44 del capítulo 1. [*Sugerencia*: construya una tabla cuadrada con las x_i en el margen izquierdo y en la parte superior. Luego calcule los promedios en la diagonal y encima de ella.]

36. Cuando la distribución de la población es normal, se puede utilizar la mediana estadística $\{|X_1 - \tilde{X}|, \ldots, |X_n - \tilde{X}|\}/.6745$ para estimar σ. Este estimador es más resistente a los efectos de los valores apartados (observaciones alejadas del grueso de los datos) que la desviación estándar muestral. Calcule tanto la estimación puntual correspondiente como s para los datos del ejemplo 6.2.

37. Cuando la desviación estándar muestral S está basada en una muestra aleatoria de una distribución de población normal, se puede demostrar que

$$E(S) = \sqrt{2/(n - 1)}\,\Gamma(n/2)\sigma/\Gamma((n - 1)/2)$$

Use ésta para obtener un estimador insesgado de σ de la forma cS. ¿Cuál es c cuando $n = 20$?

38. Cada uno de n especímenes tiene que ser pesado dos veces en la misma báscula. Sean X_i y Y_i los dos pesos observados del i-ésimo espécimen. Suponga que X_i y Y_i son independientes uno de otro, cada uno normalmente distribuido con valor medio μ_i (el peso verdadero del espécimen i) y varianza σ^2.

a. Demuestre que el estimador de probabilidad máxima de σ^2 es $\hat{\sigma}^2 = \sum(X_i - Y_i)^2/(4n)$. [*Sugerencia*: si $\bar{z} = (z_1 + z_2)/2$, entonces $\Sigma(z_i - \bar{z})^2 = (z_1 - z_2)^2/2$.]

b. ¿Es el estimador de máxima probabilidad $\hat{\sigma}^2$ un estimador insesgado de σ^2? Determine un estimador insesgado de σ^2. [*Sugerencia*: para cualquier variable aleatoria Z, $E(Z^2) = V(Z) + [E(Z)]^2$. Aplique ésta a $Z = X_i - Y_i$.]

Bibliografía

DeGroot, Morris y Mark Schervish, *Probability and Statistics* (3a. ed.), Addison-Wesley, Boston, MA, 2002. Incluye una excelente discusión tanto de propiedades generales como de métodos de estimación puntual; de particular interés son los ejemplos que muestran cómo los principios y métodos generales pueden dar estimadores insatisfactorios en situaciones particulares.

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Thomson-Brooks/Cole, Belmont, CA, 2007. La exposición es un poco más completa y compleja que la de este libro.

Efron, Bradley y Robert Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, Nueva York, 1993. La Biblia del bootstrap.

Hoaglin, David, Frederick Mosteller y John Tukey, *Understanding Robust and Exploratory Data Analysis*, Wiley, Nueva York, 1983. Contiene varios buenos capítulos sobre estimación puntual robusta, incluido uno sobre estimación *M*.

Rice, John, *Mathematical Statistics and Data Analysis* (3a. ed.), Thomson-Brooks/Cole, Belmont, CA, 2007. Una agradable mezcla de teoría y datos estadísticos.

7 Intervalos estadísticos basados en una sola muestra

INTRODUCCIÓN

Una estimación puntual, por el hecho de ser un solo número no proporciona información sobre la precisión y confiabilidad de la estimación. Considérese, por ejemplo, utilizar el estadístico $\overline{X}$ para calcular una estimación puntual de la resistencia a la ruptura promedio verdadera (g) de toallas de papel de cierta marca, y supóngase que $\bar{x} = 9322.7$. Debido a la variabilidad del muestreo, virtualmente nunca es el caso de que $\bar{x} = \mu$. La estimación puntual no dice nada sobre qué tan cerca pudiera estar de μ. Una alternativa para reportar un solo valor sensible del parámetro que se está estimando es calcular y reportar un intervalo completo de valores factibles: una *estimación de intervalo* o un *intervalo de confianza* (IC). Un intervalo de confianza siempre se calcula seleccionando primero un *nivel de confianza*, el cual mide el grado de confiabilidad del intervalo. Un intervalo de confianza con 95% de nivel de confianza de la resistencia a la ruptura promedio verdadera podría tener un límite inferior de 9162.5 y un límite superior de 9482.9. Entonces al nivel de confianza de 95%, cualquier valor de μ entre 9162.5 y 9482.5 es factible. Un nivel de confianza de 95% implica que 95% de todas las muestras daría un intervalo que incluye μ o cualquier otro parámetro que se esté estimando, y sólo 5% de las muestras darían un intervalo erróneo. Los niveles de confianza más frecuentemente utilizados son 95%, 99% y 90%. Mientras más alto es el nivel de confianza, más fuerte es la creencia de que el valor del parámetro que se está estimando queda dentro del intervalo (en breve se dará una interpretación de cualquier nivel de confianza particular).

El ancho del intervalo da información sobre la precisión de una estimación de intervalo. Si el nivel de confianza es alto y el intervalo resultante es bastante angosto, el conocimiento del valor del parámetro es razonablemente preciso. Un muy amplio intervalo de confianza, sin embargo, transmite el mensaje de que existe gran cantidad de incertidumbre sobre el valor de lo que se está estimando. La figura 7.1 muestra intervalos de confianza de 95% de resistencias a la ruptura promedio verdaderas de dos marcas diferentes de toallas de papel. Uno de estos intervalos sugiere un conocimiento preciso de μ, mientras que el otro sugiere un rango muy amplio de valores factibles.

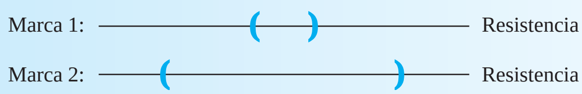


Figura 7.1 Intervalos de confianza que indican información precisa (marca 1) e imprecisa (marca 2) sobre μ

7.1 Propiedades básicas de los intervalos de confianza

Los conceptos y propiedades básicas de los intervalos de confianza son más fáciles de introducir si primero se presta atención a un problema simple, aunque un tanto irreal. Supóngase que el parámetro de interés es una media poblacional μ y que

1. La distribución de la población es normal
2. El valor de la desviación estándar σ de la población es conocido

Con frecuencia es razonable suponer que la distribución de la población es normal. Sin embargo, si el valor de μ es desconocido, no es factible que el valor de σ esté disponible (el conocimiento del centro de una población en general precede a la información con respecto a la dispersión). En las secciones 7.2 y 7.3 se desarrollarán métodos basados en suposiciones menos restrictivas.

Ejemplo 7.1 Ingenieros industriales especialistas en ergonomía se ocupan del diseño de espacios de trabajo y dispositivos operados por trabajadores con objeto de alcanzar una alta productividad y comodidad. El artículo "Studies on Ergonomically Designed Alphanumeric Keyboards" (*Human Factors*, 1985: 175–187) reporta sobre un estudio de altura preferida para un teclado experimental con un gran soporte para el antebrazo y muñeca. Se seleccionó una muestra de $n = 31$ mecanógrafos entrenados y se determinó la altura preferida del teclado de cada mecanógrafo. La altura preferida promedio muestral resultante fue de $\bar{x} = 80.0$ cm. Suponiendo que la altura preferida está normalmente distribuida con $\sigma = 2.0$ cm (un valor sugerido por los datos que aparecen en el artículo), obtenga un intervalo de confianza para μ, la altura promedio verdadera preferida por la población de todos los mecanógrafos experimentados. ■

Se supone que las observaciones muestrales reales $x_1, x_2, \ldots, x_n$ son el resultado de una muestra aleatoria $X_1, \ldots, X_n$ tomada de una distribución normal con valor medio μ y desviación estándar σ. Los resultados del capítulo 5 implican entonces que independientemente del tamaño de muestra n, la media muestral $\bar{X}$ está normalmente distribuida con valor esperado μ y desviación estándar $\sigma/\sqrt{n}$. Si se estandariza $\bar{X}$ restando primero su valor esperado y luego dividiendo entre su desviación estándar se obtiene la variable normal estándar

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{7.1}$$

Como el área bajo la curva normal estándar entre −1.96 y 1.96 es .95,

$$P\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = .95 \qquad (7.2)$$

A continuación manipúlense las desigualdades que están adentro del paréntesis en (7.2) de modo que aparezcan en la forma equivalente $l < \mu < u$, donde los puntos extremos l y u implican $\overline{X}$ y $\sigma/\sqrt{n}$. Esto se logra mediante la siguiente secuencia de operaciones, cada una de las cuales da desigualdades equivalentes a las originales.

1. Multiplíquese por $\sigma/\sqrt{n}$:

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

2. Réstese $\overline{X}$ de cada término:

$$-\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

3. Multiplíquese por −1 para eliminar el signo menos enfrente de μ (el cual invierte la dirección de cada desigualdad):

$$\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

es decir,

$$\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

La equivalencia de cada conjunto de desigualdades con el conjunto original implica que

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \qquad (7.3)$$

El evento en el interior del paréntesis en (7.3) tiene una apariencia poco común; previamente, la cantidad aleatoria aparecía a la mitad con constantes en ambos extremos, como en $a \leq Y \leq b$. En (7.3) la cantidad aleatoria aparece en los dos extremos, mientras que la constante desconocida μ aparece a la mitad. Para interpretar (7.3), considérese un **intervalo aleatorio** con el punto extremo izquierdo $\overline{X} - 1.96 \cdot \sigma/\sqrt{n}$ y punto extremo derecho $\overline{X} + 1.96 \cdot \sigma/\sqrt{n}$. En notación de intervalo, esto se transforma en

$$\left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\quad \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) \qquad (7.4)$$

El intervalo (7.4) es aleatorio porque sus dos puntos extremos implican una variable aleatoria. Está centrada en la media muestral $\overline{X}$ y se extiende a $1.96\sigma/\sqrt{n}$ a cada lado de $\overline{X}$. Por consiguiente el ancho del intervalo es $2 \cdot (1.96) \cdot \sigma/\sqrt{n}$, el cual no es aleatorio; sólo la localización del intervalo (su punto medio $\overline{X}$) lo es (figura 7.2). Ahora (7.3) puede ser parafraseado como "*la probabilidad es .95 de que el intervalo aleatorio* (7.4) *incluya o abarque el valor verdadero de* μ". Antes de realizar cualquier experimento y de recolectar cualquier dato, es bastante probable que μ estará dentro del intervalo (7.4).

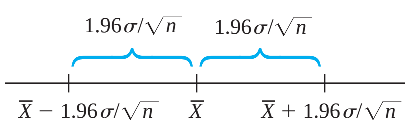


Figura 7.2 Intervalo aleatorio (7.4) con su centro en $\overline{X}$

DEFINICIÓN

Si después de observar $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, se calcula la media muestral observada $\bar{x}$ y luego se sustituye $\bar{x}$ en (7.4) en lugar de $\bar{X}$, el intervalo fijo resultante se llama **intervalo de 95% de confianza para μ.** Este intervalo de confianza se expresa como

$$\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) \text{ es un intervalo de confianza de 95\% para } \mu$$

o como

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \text{ con 95\% de confianza}$$

Una expresión concisa para el intervalo es $\bar{x} \pm 1.96 \cdot \sigma/\sqrt{n}$, donde $-$ da el punto extremo izquierdo (límite inferior) y $+$ da el punto extremo derecho (límite superior).

Ejemplo 7.2 (Continuación del ejemplo 7.1)

Las cantidades requeridas para calcular el intervalo de confianza de 95% para la altura preferida promedio verdadera son $\sigma = 2.0$, $n = 31$ y $\bar{x} = 80.0$. El intervalo resultante es

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 80.0 \pm (1.96)\frac{2.0}{\sqrt{31}} = 80.0 \pm .7 = (79.3, 80.7)$$

Es decir, se puede estar totalmente confiado, en el nivel de confianza de 95%, de que $79.3 < \mu < 80.7$. Este intervalo es relativamente angosto, lo que indica que μ ha sido estimada con bastante precisión. ■

Interpretación de un intervalo de confianza

El nivel de confianza de 95% para el intervalo que se acaba de definir fue heredado del .95 de probabilidad para el intervalo aleatorio (7.4). Los intervalos con otros niveles de confianza serán introducidos en breve. Por ahora, más bien, considérese cómo se puede interpretar el 95% de confianza.

Como se inició con un evento cuya probabilidad era de .95 —que el intervalo aleatorio (7.4) capturaría el valor verdadero de μ—, y luego se utilizaron los datos del ejemplo 7.1 para calcular el intervalo de confianza (79.3, 80.7), es tentador concluir que μ está dentro de este intervalo fijo con probabilidad de .95. Pero al sustituir $\bar{x} = 80.0$ en lugar de $\bar{X}$, toda la aleatoriedad desaparece; el intervalo (79.3, 80.7) no es un intervalo aleatorio y μ es una constante (desafortunadamente desconocida). Es por consiguiente *incorrecto* escribir la proposición $P(\mu$ quede en $(79.3, 80.7)) = .95$.

Una interpretación correcta de "95% de confianza" se basa en la interpretación de probabilidad de frecuencia relativa a largo plazo: decir que un evento A tiene una probabilidad de .95 es decir que si el experimento en el cual se definió A se realiza una y otra vez, a la larga A ocurrirá el 95% del tiempo. Supóngase que se obtiene otra muestra de alturas preferidas por los mecanógrafos y se calcula otro intervalo de 95%. Luego se considera repetir esto con una tercera muestra, una cuarta, una quinta, y así sucesivamente. Sea A el evento en que $\bar{X} - 1.96 \cdot \sigma/\sqrt{n} < \mu < \bar{X} + 1.96 \cdot \sigma/\sqrt{n}$. Ya que $P(A) = .95$, a la larga el 95% de los intervalos de confianza calculados contendrán a μ. Esto se ilustra en la figura 7.3, donde la línea vertical corta el eje de medición en el valor verdadero (pero desconocido) de μ. Observe que 7 de los 100 intervalos mostrados fallan al contener a μ. A la larga, sólo 5% de los intervalos construidos así no contendrán a μ.

De acuerdo con esta interpretación, el nivel de confianza de 95% no es en sí una proposición sobre cualquier intervalo particular tal como (79.3, 80.7). En su lugar pertenece

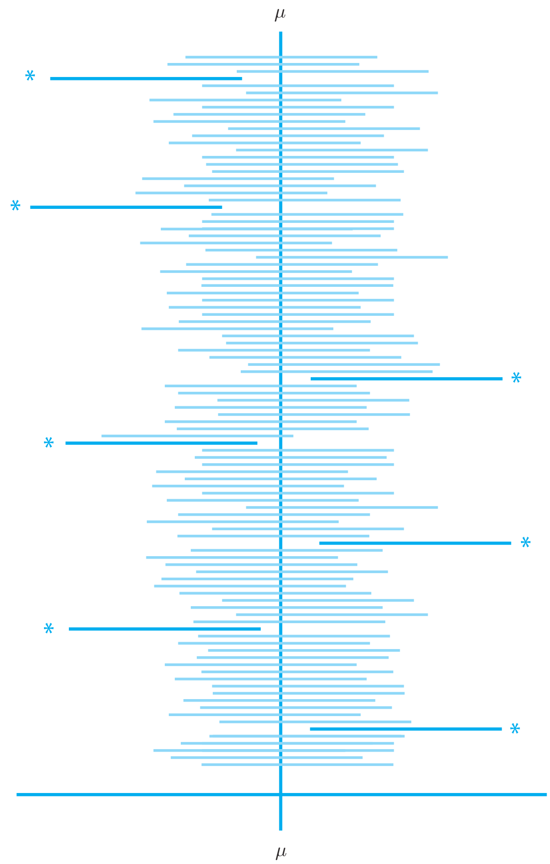


Figura 7.3 Cien niveles de confianza de 95% (los asteriscos identifican intervalos que no incluyen a μ).

a lo que sucedería si se construyera un número muy grande de intervalos parecidos por medio de la misma fórmula de intervalo de confianza. Aunque esto puede parecer no satisfactorio, el origen de la dificultad yace en la interpretación de probabilidad, es válida para una larga secuencia de réplicas de un experimento en lugar de sólo para una. Existe otro método para abordar la construcción e interpretación de intervalos de confianza que utiliza la noción de probabilidad subjetiva y el teorema de probabilidad de Bayes, aunque los detalles técnicos se salen del alcance de este libro; el libro de DeGroot y colaboradores (véase la bibliografía del capítulo 6) es una buena fuente. El intervalo presentado aquí (así como también cada intervalo presentado subsecuentemente) se llama intervalo de confianza "clásico" porque su interpretación se apoya en la noción clásica de probabilidad.

Otros niveles de confianza

El nivel de confianza de 95% fue heredado de la probabilidad de .95 de las desigualdades iniciales que aparecen en (7.2). Si se desea un nivel de confianza de 99%, la probabilidad inicial de .95 debe ser reemplazada por .99, lo que implica cambiar el valor crítico z de 1.96 a 2.58. Un intervalo de confianza de 99% resulta entonces de utilizar 2.58 en lugar de 1.96 en la fórmula para el intervalo de confianza de 95%.

De hecho, cualquier nivel de confianza deseado se obtiene reemplazando 1.96 o 2.58 con el valor crítico normal estándar apropiado. Como la figura 7.4 muestra, utilizando $z_{\alpha/2}$ en lugar de 1.96 se logra una probabilidad de $1 - \alpha$.

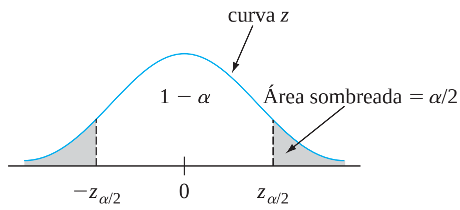


Figura 7.4 $P(-z_{\alpha/2} \leq Z < z_{\alpha/2}) = 1 - \alpha$

DEFINICIÓN

La siguiente expresión da un **intervalo de confianza de 100(1 − α)%** para la media μ de una población normal cuando se conoce el valor de σ

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{7.5}$$

o, de forma equivalente, por $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

La fórmula (7.5) para el intervalo de confianza también se puede expresar en palabras como

estimación puntual de μ ± (valor crítico z) (error estándar de la media).

Ejemplo 7.3 No hace mucho tiempo que el proceso de producción de una caja de control de un tipo particular para un motor fue modificado. Antes de esta modificación, datos históricos sugirieron que la distribución de los diámetros de agujeros para bujes en las cajas era normal con desviación estándar de .100 mm. Se cree que la modificación no ha afectado la forma de la distribución o la desviación estándar, pero que el valor del diámetro medio pudo haber cambiado. Se selecciona una muestra de 40 cajas y se determina el diámetro de agujero para cada una, y el resultado es un diámetro medio muestral de 5.426 mm. Calcúlese un intervalo de confianza para el diámetro de agujero promedio verdadero utilizando un nivel de confianza de 90%. Esto requiere que $100(1 - \alpha) = 90$, de donde $\alpha = .10$ y $z_{\alpha/2} = z_{.05} = 1.645$ (correspondiente a un área de curva z acumulativa de .9500). El intervalo deseado es entonces

$$5.426 \pm (1.645)\frac{.100}{\sqrt{40}} = 5.426 \pm .026 = (5.400, 5.452)$$

Con un razonablemente alto grado de confianza, se puede decir que $5.400 < \mu < 5.452$. Este intervalo es algo angosto debido a la pequeña cantidad de variabilidad del diámetro del agujero ($\sigma = .100$). ■

Nivel de confianza, precisión y tamaño de muestra

¿Por qué decidirse por un nivel de confianza de 95% cuando un nivel de 99% es alcanzable? Porque el precio pagado por el nivel de confianza más alto es un intervalo más ancho. Como el intervalo de 95% se extiende $1.96 \cdot \sigma/\sqrt{n}$ a cada lado de $\bar{x}$, el ancho del intervalo es $2(1.96) \cdot \sigma/\sqrt{n} = 3.92 \cdot \sigma/\sqrt{n}$. Asimismo, el ancho del intervalo de 99% es $2(2.58) \cdot \sigma/\sqrt{n} = 5.16 \cdot \sigma/\sqrt{n}$. Es decir, se tiene más confianza en el intervalo de 99% precisamente porque es más ancho. Mientras más alto es el grado de confianza deseado, más ancho es el intervalo resultante.

Si se considera que el ancho del intervalo especifica su precisión o exactitud, entonces el nivel de confianza (o confiabilidad) del intervalo está relacionado de manera inversa con su precisión. La estimación de un intervalo altamente confiable puede ser imprecisa por el hecho de que los puntos extremos del intervalo pueden estar muy alejados, mientras que un intervalo preciso puede acarrear una confiabilidad relativamente baja. Por consiguiente no se puede decir de modo inequívoco que se tiene que preferir un intervalo de 99% a uno de 95%; la ganancia de confiabilidad acarrea una pérdida de precisión.

Una estrategia atractiva es especificar tanto el nivel de confianza deseado como el ancho del intervalo y luego determinar el tamaño de muestra necesario.

Ejemplo 7.4 Un monitoreo exhaustivo de un sistema de tiempo compartido de computadoras sugiere que el tiempo de respuesta a un comando de edición particular está normalmente distribuido con desviación estándar de 25 milisegundos. Se instaló un nuevo sistema operativo y se desea estimar el tiempo de respuesta promedio verdadero μ en el nuevo entorno. Suponiendo que los tiempos de respuesta siguen estando normalmente distribuidos con $\sigma = 25$, ¿qué tamaño de muestra es necesario para asegurarse de que el intervalo de confianza de 95% resultante tiene un ancho de (cuando mucho) 10? El tamaño de muestra n debe satisfacer

$$10 = 2 \cdot (1.96)(25/\sqrt{n})$$

Reordenando esta ecuación se obtiene

$$\sqrt{n} = 2 \cdot (1.96)(25)/10 = 9.80$$

por consiguiente

$$n = (9.80)^2 = 96.04$$

En vista de que n debe ser un entero, se requiere un tamaño de muestra de 97. ■

Una fórmula general para el tamaño de muestra n necesario para garantizar un ancho de intervalo w se obtiene igualando w a $2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y despejando n.

> El tamaño de muestra necesario para que el intervalo de confianza (7.5) dé un ancho w es
>
> $$n = \left(2z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Mientras más pequeño es el ancho deseado w, más grande debe ser n. Además, n es una función creciente de σ (más variabilidad de la población requiere un tamaño de muestra más grande) y del nivel de confianza $100(1 - \alpha)$ (conforme α decrece, $z_{\alpha/2}$ se incrementa).

La mitad del ancho $1.96\sigma/\sqrt{n}$ del intervalo de confianza de 95% en ocasiones se llama **límite en el error de estimación** asociado con un nivel de confianza de 95%. Es decir, con 95% de confianza, la estimación puntual $\bar{x}$ no estará a más de esta distancia de μ. Antes de obtener datos, es posible que un investigador desee determinar un tamaño de muestra con el cual se logre un valor particular del límite. Por ejemplo, si μ representa la eficiencia de combustible promedio (mpg) de todos los carros de cierto tipo, el objetivo de una investigación puede ser estimar μ dentro de 1 mpg con 95% de confianza. Más generalmente, si se desea estimar μ dentro de una cantidad B (el límite especificado en el error de estimación) con confianza de $100(1 - \alpha)$ %, el tamaño de muestra necesario se obtiene al reemplazar $2/w$ por $1/B$ en el recuadro precedente.

Deducción de un intervalo de confianza

Sean $X_1, X_2, \ldots, X_n$ la muestra en la cual se tiene que basar el intervalo de confianza para un parámetro θ. Supóngase que se puede determinar una variable aleatoria que satisfaga las dos siguientes propiedades:

1. La variable depende funcionalmente tanto de $X_1, \dots, X_n$ como de θ.
2. La distribución de probabilidad de la variable no depende de θ ni de cualesquiera otros parámetros desconocidos.

Sea $h(X_1, X_2, \dots, X_n; \theta)$ esta variable aleatoria. Por ejemplo, si la distribución de la población es normal con σ y $\theta = \mu$ conocidos, la variable $h(X_1, \dots, X_n; \mu) = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ satisface ambas propiedades; claramente depende funcionalmente de μ, no obstante su distribución de probabilidad normal estándar, la cual no depende de μ. En general, la forma de la función h casi siempre se pone de manifiesto al examinar la distribución de un estimador apropiado $\hat{\theta}$.

Con cualquier α entre 0 y 1, se ve que las constantes a y b satisfacen

$$P(a < h(X_1, \dots, X_n; \theta) < b) = 1 - \alpha \tag{7.6}$$

A causa de la segunda propiedad, a y b no dependen de θ. En el ejemplo normal, $a = -z_{\alpha/2}$ y $b = z_{\alpha/2}$. Ahora supóngase que las desigualdades en (7.6) pueden ser manipuladas para aislar θ, y así obtener la proposición de probabilidad equivalente

$$P(l(X_1, X_2, \dots, X_n) < \theta < u(X_1, X_2, \dots, X_n)) = 1 - \alpha$$

Entonces $l(x_1, x_2, \dots, x_n)$ y $u(x_1, \dots, x_n)$ son los límites de confianza inferior y superior, respectivamente, para un intervalo de confianza de $100(1 - \alpha)\%$. En el ejemplo normal, se vio que $l(X_1, \dots, X_n) = \overline{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y $u(X_1, \dots, X_n) = \overline{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Ejemplo 7.5 Un modelo teórico sugiere que el tiempo hasta la ruptura de un fluido aislante entre electrodos a un voltaje particular tiene una distribución exponencial con parámetro λ (véase la sección 4.4). Una muestra aleatoria de $n = 10$ tiempos de ruptura da los siguientes datos muestrales (en min): $x_1 = 41.53$, $x_2 = 18.73$, $x_3 = 2.99$, $x_4 = 30.34$, $x_5 = 12.33$, $x_6 = 117.52$, $x_7 = 73.02$, $x_8 = 223.63$, $x_9 = 4.00$, $x_{10} = 26.78$. Se desea un intervalo de confianza de 95% para λ y para el tiempo de ruptura promedio verdadero.

Sea $h(X_1, X_2, \dots, X_n; \lambda) = 2\lambda\sum X_i$. Se puede demostrar que esta variable aleatoria tiene una distribución de probabilidad llamada distribución ji cuadrada con $2n$ grados de libertad (gl) ($\nu = 2n$, donde ν es el parámetro de una distribución ji cuadrada como se menciona en la sección 4.4). La tabla A.7 del apéndice ilustra una curva de densidad ji cuadrada típica y tabula valores críticos que capturan áreas de colas específicas. El número pertinente de grados de libertad en este caso es $2(10) = 20$. La fila $\nu = 20$ de la tabla muestra que 34.170 captura un área de cola superior de .025 y 9.591 captura un área de cola inferior de .025 (área de cola superior de .975). Por consiguiente con $n = 10$,

$$P(9.591 < 2\lambda\sum X_i < 34.170) = .95$$

La división entre $2\sum X_i$ aísla λ y se obtiene

$$P(9.591/(2\sum X_i) < \lambda < (34.170/(2\sum X_i)) = .95$$

El límite inferior del intervalo de confianza de 95% para λ es $9.591/(2\sum x_i)$, y el límite superior es $34.170/(2\sum x_i)$. Con los datos dados $\sum x_i = 550.87$ da el intervalo (.00871, .03101).

El valor esperado de una variable aleatoria exponencial es $\mu = 1/\lambda$. Puesto que

$$P(2\sum X_i/34.170 < 1/\lambda < 2\sum X_i/9.591) = .95$$

el intervalo de confianza de 95% para el tiempo de ruptura promedio verdadero es $(2\sum x_i/34.170, 2\sum x_i/9.591) = (32.24, 114.87)$. Obviamente este intervalo es bastante ancho, lo que refleja una variabilidad sustancial de los tiempos de ruptura y un pequeño tamaño de muestra. ■

En general, los límites de confianza superior e inferior resultan de reemplazar cada $<$ en (7.6) por $=$ y resolviendo para θ. En el ejemplo del fluido aislante que se acaba de considerar, $2\lambda\sum x_i = 34.170$ da $\lambda = 34.170/(2\sum x_i)$ como límite de confianza superior y el límite inferior se obtiene con la otra ecuación. Obsérvese que los dos límites de intervalo no están equidistantes de la estimación puntual, en vista de que el intervalo no es de la forma $\hat{\theta} \pm c$.

Intervalos de confianza bootstrap

La técnica bootstrap se introdujo en el capítulo 6 como una forma de estimar $\sigma_{\hat{\theta}}$. También puede ser aplicada para obtener un intervalo de confianza para θ. Considérese de nuevo la estimación de la media μ de una distribución normal cuando σ es conocido. Reemplácese μ con θ y úsese $\hat{\theta} = \overline{X}$ como estimador puntual. Obsérvese que $1.96\sigma\sqrt{n}$ es el 97.5º percentil de la distribución de $\hat{\theta} - \theta$ [esto es, $P(\overline{X} - \mu < 1.96\sigma/\sqrt{n}) = P(Z < 1.96) = .9750$]. Del mismo modo, $-1.96\sigma/\sqrt{n}$ es el 2.5º percentil, por consiguiente

$$\begin{aligned} .95 &= P(2.5^{\text{o}} \text{ percentil} < \hat{\theta} - \theta < 97.5^{\text{o}} \text{ percentil}) \\ &= P(\hat{\theta} - 2.5^{\text{o}} \text{ percentil} > \theta > \hat{\theta} - 97.5^{\text{o}} \text{ percentil}) \end{aligned}$$

Es decir, con

$$\begin{aligned} l &= \hat{\theta} - 97.5^{\text{o}} \text{ percentil de } \hat{\theta} - \theta \\ u &= \hat{\theta} - 2.5^{\text{o}} \text{ percentil de } \hat{\theta} - \theta \end{aligned} \tag{7.7}$$

el intervalo de confianza para θ es (l, u). En muchos casos, los percentiles en (7.7) no pueden ser calculados, pero sí *pueden* ser estimados con muestras bootstrap. Supóngase que se obtienen $B = 1000$ muestras bootstrap y se calculan $\hat{\theta}_1^*, \ldots, \hat{\theta}_{1000}^*$ y $\bar{\theta}^*$ seguidos por las 1000 diferencias $\hat{\theta}_1^* - \bar{\theta}^*, \ldots, \hat{\theta}_{1000}^* - \bar{\theta}^*$. La 25º más grande y la 25º más pequeña de estas diferencias son estimaciones de los percentiles desconocidos en (7.7). Consúltese los libros de Devore y Berk o de Efron citados en el capítulo 6 para más información.

EJERCICIOS Sección 7.1 (1–11)

1. Considere una distribución de población normal con el valor de σ conocido.

a. ¿Cuál es el nivel de confianza para el intervalo $\bar{x} \pm 2.81\sigma/\sqrt{n}$?

b. ¿Cuál es el nivel de confianza para el intervalo $\bar{x} \pm 1.44\sigma/\sqrt{n}$?

c. ¿Qué valor de $z_{\alpha/2}$ en la fórmula de intervalo de confianza (7.5) da un nivel de confianza de 99.7%?

d. Responda la pregunta hecha en el inciso (c) para un nivel de confianza de 75%.

2. Cada uno de los siguientes es un intervalo de confianza para μ = frecuencia de resonancia promedio (es decir, media de la población) verdadera (Hz) para todas las raquetas de tenis de un tipo:

(114.4, 115.6) (114.1, 115.9)

a. ¿Cuál es el valor de la frecuencia de resonancia media muestral?

b. Ambos intervalos se calcularon con los mismos datos muestrales. El nivel de confianza para uno de estos intervalos es de 90% y para el otro es de 99%. ¿Cuál de los intervalos tiene el nivel de confianza de 90% y por qué?

3. Suponga que se selecciona una muestra aleatoria de 50 botellas de una marca particular de jarabe para la tos y se determina el contenido de alcohol de cada una. Sea μ el contenido promedio de alcohol de la población de todas las botellas de la marca estudiada. Suponga que el intervalo de confianza de 95% resultante es (7.8, 9.4).

a. ¿Un intervalo de confianza de 90% calculado con esta muestra habría resultado más angosto o más ancho que el intervalo dado? Explique su razonamiento.

b. Considere la siguiente proposición: existe 95% de probabilidades de que μ esté entre 7.8 y 9.4. ¿Es correcta esta proposición? ¿ Por qué sí o por qué no?

c. Considere la siguiente proposición: se puede estar totalmente confiado de que 95% de todas las botellas de este tipo de jarabe para la tos tienen un contenido de alcohol de entre 7.8 y 9.4. ¿Es correcta esta proposición? ¿Por qué sí o por qué no?

d. Considere la siguiente proposición: si el proceso de selección de una muestra de tamaño 50 y el cálculo del intervalo de 95% correspondiente se repite 100 veces, 95 de los intervalos resultantes incluirán μ. ¿Es correcta esta proposición? ¿Por qué sí o por qué no?

4. Se desea un intervalo de confianza para la pérdida por carga parásita promedio verdadera μ (watts) de cierto tipo de motor de inducción cuando la corriente a través de la línea se mantiene en 10 amps a una velocidad de 1500 rpm. Suponga que la pérdida por carga parásita está normalmente distribuida con $\sigma = 3.0$.

a. Calcule un intervalo de confianza de 95% para μ cuando $n = 25$ y $\bar{x} = 58.3$.

b. Calcule un intervalo de confianza de 95% para μ cuando $n = 100$ y $\bar{x} = 58.3$.

c. Calcule un intervalo de confianza de 99% para μ cuando $n = 100$ y $\bar{x} = 58.3$.

d. Calcule un intervalo de confianza de 82% para μ cuando $n = 100$ y $\bar{x} = 58.3$.

e. ¿Qué tan grande debe ser n si el ancho del intervalo de 99% para μ tiene que ser 1.0?

5. Suponga que la porosidad al helio (en porcentaje) de muestras de carbón tomadas de cualquier costura particular está normalmente distribuida con desviación estándar verdadera de .75.

a. Calcule un intervalo de confianza de 95% para la porosidad promedio verdadera de una costura si la porosidad promedio en 20 especímenes de la costura fue de 4.85.

b. Calcule un intervalo de confianza de 98% para la porosidad promedio verdadera de otra costura basada en 16 especímenes con porosidad promedio muestral de 4.56.

c. ¿Qué tan grande debe ser un tamaño de muestra si el ancho del intervalo de 95% tiene que ser de .40?

d. ¿Qué tamaño de muestra se necesita para estimar la porosidad promedio verdadera dentro de .2 con confianza de 99%?

6. Con base en pruebas extensas, se sabe que el punto de cedencia de un tipo particular de varilla de refuerzo de acero suave está normalmente distribuido con $\sigma = 100$. La composición de la varilla se modificó un poco, pero no se cree que la alteración haya afectado la normalidad o el valor de σ.

a. Suponiendo que éste sea el caso, si una muestra de 25 varillas modificadas dio por resultado un punto de cedencia promedio muestral de 8439 lb, calcule un intervalo de confianza de 90% para el punto de cedencia promedio verdadero de la varilla modificada.

b. ¿Cómo modificaría el intervalo del inciso (a) para obtener un nivel de confianza de 92%?

7. ¿En cuánto se debe incrementar el tamaño de muestra n si el ancho del intervalo de confianza (7.5) tiene que ser reducido a la mitad? Si el tamaño de muestra se incrementa por un factor de 25, ¿qué efecto tendrá esto en el ancho del intervalo? Justifique sus aseveraciones.

8. Sea $\alpha_1 > 0$, $\alpha_2 > 0$, con $\alpha_1 + \alpha_2 = \alpha$. Entonces

$$P\left(-z_{\alpha_1} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha_2}\right) = 1 - \alpha$$

a. Use esta ecuación para obtener una expresión más general para un intervalo de confianza de $100(1 - \alpha)\%$ para μ del cual el intervalo (7.5) es un caso especial.

b. Sea $\alpha = .05$ y $\alpha_1 = \alpha/4$, $\alpha_2 = 3\alpha/4$. ¿Esto resulta en un intervalo más angosto o más ancho que el intervalo (7.5)?

9. a. En las mismas condiciones que aquellas que conducen al intervalo (7.5), $P[(\bar{X} - \mu)/(\sigma/\sqrt{n}) < 1.645] = .95$. Use esta expresión para deducir un intervalo unilateral para μ de ancho infinito y que proporcione un límite de confianza inferior para μ. ¿Cuál es el intervalo para los datos del ejercicio 5(a)?

b. Generalice el resultado del inciso (a) para obtener un límite inferior con nivel de confianza de $100(1 - \alpha)\%$.

c. ¿Cuál es un intervalo análogo al del inciso (b) que proporcione un límite superior para μ? Calcule este intervalo de 99% para los datos del ejercicio 4(a).

10. Una muestra aleatoria de $n = 15$ bombas térmicas de cierto tipo produjo las siguientes observaciones de vida útil (en años):

2.0	1.3	6.0	1.9	5.1	.4	1.0	5.3
15.7	.7	4.8	.9	12.2	5.3	.6	

a. Suponga que la distribución de la vida útil es exponencial y use un argumento paralelo al del ejemplo 7.5 para obtener un intervalo de confianza de 95% para la vida útil esperada (promedio verdadero).

b. ¿Cómo debería modificarse el intervalo del inciso (a) para obtener un nivel de confianza de 99%?

c. ¿Cuál es un intervalo de confianza de 95% para la desviación estándar de la distribución de la vida útil? [*Sugerencia*: ¿cuál es la desviación estándar de una variable aleatoria exponencial?]

11. Considere los siguientes 1000 intervalos de confianza de 95% para μ que un consultor estadístico obtendrá para varios clientes. Suponga que se seleccionan independientemente uno de otro los conjuntos de datos en los cuales están basados los intervalos. ¿Cuántos de estos 1000 intervalos espera que capturen el valor correspondiente de μ? ¿Cuál es la probabilidad de que entre 940 y 960 de estos intervalos contengan el valor correspondiente de μ? [*Sugerencia*: sea Y = el número entre los 1000 intervalos que contienen a μ. ¿Qué clase de variable aleatoria es Y?]

7.2 Intervalos de confianza de muestra grande para una media y proporción de población

Se supuso en el intervalo de confianza para μ dado en la sección previa que la distribución de la población es normal con el valor de σ conocido. A continuación se presenta un intervalo de confianza de muestra grande cuya validez no requiere estas suposiciones. Después de demostrar cómo el argumento que lleva a este intervalo se aplica en forma extensa para producir otros intervalos de muestra grande, habrá que enfocarse en un intervalo para una proporción de población p.

Intervalo de muestra grande para μ

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con media μ y desviación estándar σ. Siempre que n es grande, el teorema del límite central implica que $\overline{X}$ tiene de manera aproximada una distribución normal cualquiera que sea la naturaleza de la distribución de la población. Se deduce entonces que $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ tiene aproximadamente una distribución estándar normal, de modo que

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Un argumento paralelo al dado en la sección 7.1 da $\overline{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ como intervalo de confianza de muestra grande para μ con un nivel de confianza de *aproximadamente* $100(1 - \alpha)\%$. Es decir, cuando n es grande, el intervalo de confianza para μ dado antes permanece válido cualquiera que sea la distribución de la población, siempre que el calificador esté insertado "aproximadamente" enfrente del nivel de confianza.

Una dificultad práctica con este desarrollo es que el cálculo del intervalo de confianza requiere el valor de σ, el cual rara vez es conocido. Considérese la variable estandarizada $(\overline{X} - \mu)/(S/\sqrt{n})$, en la cual la desviación estándar muestral S ha reemplazado a σ. Previamente había aleatoriedad sólo en el numerador de Z gracias a $\overline{X}$. En la nueva variable estandarizada, tanto $\overline{X}$ *como* S cambian de valor de una muestra a otra. Así que aparentemente la distribución de la nueva variable deberá estar más dispersa que la curva z para reflejar la variación extra en el denominador. Esto en realidad es cierto cuando n es pequeño. Sin embargo, con n grande la sustitución de S en lugar de σ agrega un poco de variabilidad extra, así que esta variable también tiene aproximadamente una distribución normal estándar. La manipulación de la variable en la proposición de probabilidad, como en el caso de σ conocida, da un intervalo de confianza de muestra grande general para μ.

PROPOSICIÓN

Si n es suficientemente grande, la variable estandarizada

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene aproximadamente una distribución normal estándar. Esto implica que

$$\overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \tag{7.8}$$

es un **intervalo de confianza de muestra grande para μ** con nivel de confianza aproximadamente de $100(1 - \alpha)\%$. Esta fórmula es válida sin importar la forma de la distribución de la población.

Es decir, el intervalo de confianza (7.8) es la

estimación puntual de $\mu \pm$ (z valor crítico) (error estándar estimado de la media).

En general, $n > 40$ será suficiente para justificar el uso de este intervalo. Esto es algo más conservador que la regla empírica del teorema del límite central debido a la variabilidad adicional introducida por el uso de S en lugar de σ.

Ejemplo 7.6 ¿Siempre quiso tener un Porsche? El autor pensó que tal vez podía permitirse un Boxster, el modelo más barato. Así que se fue a www.cars.com el 18 de noviembre de 2009 y encontró un total de 1113 automóviles de este tipo en la lista. Preguntando, los precios iban desde \$ 3499 a \$ 130,000 (el precio de este último fue uno de los dos que excedían los

$ 70,000). Los precios lo deprimieron, por lo que en cambio se centró en las lecturas del odómetro (millas). Aquí se presentan las lecturas de una muestra de 50 de estos Boxster:

2948	2996	7197	8338	8500	8759	12710	12925
15767	20000	23247	24863	26000	26210	30552	30600
35700	36466	40316	40596	41021	41234	43000	44607
45000	45027	45442	46963	47978	49518	52000	53334
54208	56062	57000	57365	60020	60265	60803	62851
64404	72140	74594	79308	79500	80000	80000	84000
113000	118634						

Una gráfica de caja de los datos (figura 7.5) muestra que, a excepción de los dos valores límite en el extremo superior, la distribución de valores es bastante simétrica (de hecho, una gráfica de probabilidad normal muestra un patrón bastante lineal, aunque los puntos correspondientes a las dos observaciones más pequeñas y a las dos mayores están un tanto alejadas de un ajuste lineal a través de los puntos restantes).

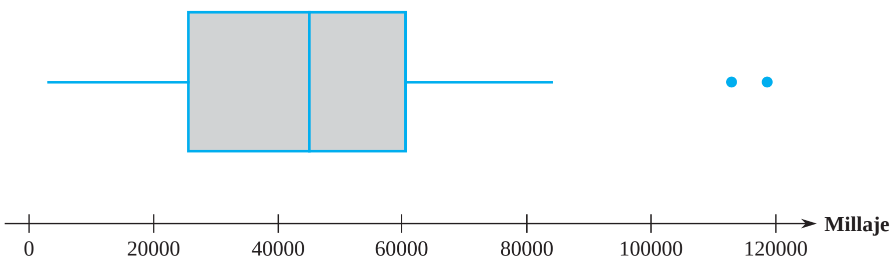


Figura 7.5 Diagrama de caja para las lecturas del odómetro del ejemplo 7.6

Las cantidades resumidas incluyen $n = 50$, $\bar{x} = 45{,}679.4$, $\tilde{x} = 45{,}013.5$, $s = 26{,}641.675$, $f_s = 34{,}265$. La media y la mediana están relativamente cerca (si los dos valores mayores fueran reducidos por 30,000, la media bajaría a 44,479.4, mientras que la mediana no se vería afectada). El diagrama de caja y las magnitudes de s y f_s respecto a la media y la mediana de ambos indican una cantidad considerable de variabilidad. El intervalo de confianza de 95% requiere que $z_{.025} = 1.96$, y el intervalo es entonces

$$45{,}679.4 \pm (1.96)\left(\frac{26{,}641.675}{\sqrt{50}}\right) = 45{,}679.4 \pm 7384.7$$
$$= (38{,}294.7, 53{,}064.1)$$

Es decir, $38{,}294.7 < \mu < 53{,}064.1$ con un nivel de confianza de aproximadamente 95%. Este intervalo es bastante amplio debido a un tamaño de muestra de 50, que aunque es grande por nuestra regla general, no es lo suficientemente grande como para superar la variabilidad en la muestra. No tenemos una estimación muy precisa de la población media de la lectura del odómetro.

¿Es el intervalo que hemos calculado uno de los 95% que en el largo plazo incluyen el parámetro calculado o es uno de los "malos" del 5% que no lo hace? Sin saber el valor de μ, no podemos decir. Recuerde que el nivel de confianza se refiere al porcentaje de captura a largo plazo cuando la fórmula se utiliza repetidamente en varias muestras; no se puede interpretar para una sola muestra y el intervalo resultante. ■

Desafortunadamente, la selección del tamaño de muestra para que dé un ancho de intervalo deseado no es simple en este caso como lo fue en el caso de σ conocida. Por eso el ancho de (7.8) es $2z_{\alpha/2}s/\sqrt{n}$. Como el valor de s no está disponible antes de que los datos hayan sido recopilados, el ancho del intervalo no puede ser determinado tan sólo con la selección de n. La única opción para un investigador que desea especificar un ancho deseado es hacer una suposición educada de cuál podría ser el valor de s. Siendo conservador y suponiendo un valor más grande de s, se seleccionará un n más grande de lo necesario. El investigador puede ser capaz de especificar un valor razonablemente preciso del rango de población (la diferencia entre los valores más grande y más pequeño). Entonces si la distribución de la población no es demasiado asimétrica, si se divide el rango entre 4 se obtiene un valor aproximado de lo que s podría ser.

Ejemplo 7.7 El tiempo de carga (minutos) para el acero de carbono en un tipo de horno de hogar abierto se determinará para cada calor en una muestra de tamaño n. Si el investigador cree que casi todos los tiempos en la distribución están entre 320 y 440, ¿qué tamaño de la muestra sería apropiado para estimar el tiempo promedio real a cuando mucho 5 minutos con un nivel de confianza del 95%?

Un valor razonable para s es $(440 - 320)/4 = 30$. Por tanto

$$n = \left[\frac{(1.96)(30)}{5}\right]^2 = 138.3$$

Dado que el tamaño de la muestra debe ser un número entero, $n = 139$ debe ser utilizado. Tenga en cuenta que la estimación está dentro de 5 minutos con el nivel de confianza especificado que es equivalente a un ancho de intervalo de confianza de 10 minutos. ■

Un intervalo de confianza de muestra grande general

Los intervalos de muestra grande $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ son casos especiales de un intervalo de confianza de muestra grande general para un parámetro θ. Suponga que $\hat{\theta}$ es un estimador que satisface las siguientes propiedades: (1) Tiene aproximadamente una distribución normal; (2) es insesgado (por lo menos aproximadamente); y (3) una expresión para $\sigma_{\hat{\theta}}$, la desviación estándar de $\hat{\theta}$, está disponible. Por ejemplo, en el caso $\theta = \mu$, $\hat{\mu} = \overline{X}$ es un estimador insesgado cuya distribución es aproximadamente normal cuando n es grande y $\sigma_{\hat{\mu}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$. Estandarizando $\hat{\theta}$ se obtiene la variable aleatoria $Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$, la cual tiene aproximadamente una distribución normal estándar. Esto justifica la proposición de probabilidad

$$P\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) \approx 1 - \alpha \qquad (7.9)$$

Suponga, primero, que $\sigma_{\hat{\theta}}$ no involucra ningún parámetro desconocido (p. ej., σ conocida en el caso $\theta = \mu$). Entonces si se reemplaza cada $<$ en (7.9) por $=$ se obtiene $\theta = \hat{\theta} \pm z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, por consiguiente los límites de confianza inferior y superior son $\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$ y $\hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, respectivamente. Suponga ahora que $\sigma_{\hat{\theta}}$ no implica a θ pero sí implica por lo menos otro parámetro desconocido. Sea $s_{\hat{\theta}}$ la estimación de $\sigma_{\hat{\theta}}$ obtenida utilizando estimaciones en lugar de los parámetros desconocidos (p. ej., $s/\sqrt{n}$ estima $\sigma/\sqrt{n}$). En condiciones generales (esencialmente que $s_{\hat{\theta}}$ se aproxime a $\sigma_{\hat{\theta}}$ con la mayoría de las muestras), un intervalo de confianza válido es $\hat{\theta} \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$. El intervalo muestral grande $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ es un ejemplo.

Por último, suponga que $\sigma_{\hat{\theta}}$ implica el θ desconocido. Éste es el caso, por ejemplo, cuando $\theta = p$, una proporción de la población. Entonces $(\hat{\theta} - \theta)/\sigma_{\hat{\theta}} = z_{\alpha/2}$ puede ser difícil de resolver. Con frecuencia se puede obtener una solución aproximada reemplazando θ en $\sigma_{\hat{\theta}}$ por su estimación $\hat{\theta}$. Esto da una desviación estándar estimada $s_{\hat{\theta}}$ y el intervalo correspondiente es de nuevo $\hat{\theta} \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$.

Es decir, este intervalo de confianza es una

estimación puntual de θ $\pm$ (valor crítico z) (error estándar estimado del estimador).

Un intervalo de confianza para una proporción de población

Sea p la proporción de "éxitos" en una población, donde *éxito* identifica a un individuo u objeto que tiene una propiedad específica (p. ej., individuos que se graduaron en una universidad, computadoras que no requieren servicio de garantía, etc.). Una muestra aleatoria de n individuos tiene que ser seleccionada y X es el número de éxitos en la muestra. Siempre que n sea pequeño comparado con el tamaño de la población, X puede ser considerada como una variable aleatoria binomial con $E(X) = np$ y $\sigma_X = \sqrt{np(1-p)}$. Además, si tanto $np \geq 10$ como $nq \geq 10$, $(q = 1 - p)$, X tiene aproximadamente una distribución normal.

El estimador natural de p es $\hat{p} = X/n$, la fracción muestral de éxitos. Como $\hat{p}$ es simplemente X multiplicada por la constante $1/n$, $\hat{p}$ también tiene aproximadamente una distribución normal. Como se muestra en la sección 6.1, $E(\hat{p}) = p$ (insesgado) y $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. La desviación estándar $\sigma_{\hat{p}}$ implica el parámetro desconocido p. Si se estandariza $\hat{p}$ restando p y dividiendo entre $\sigma_{\hat{p}}$ entonces se tiene

$$P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Procediendo como se sugirió en la subsección "Deducción de un intervalo de confianza" (sección 7.1), los límites de confianza se obtienen al reemplazar cada $<$ por $=$ y resolver la ecuación cuadrática resultante para p. Esto da las dos raíces

$$\begin{aligned} p &= \frac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n} \\ &= \tilde{p} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n} \end{aligned}$$

PROPOSICIÓN

Sea $\tilde{p} = \dfrac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n}$. Entonces, un **intervalo de confianza para una proporción de población p** con nivel de confianza aproximadamente de $100(1 - \alpha)\%$ tiene

$$\tilde{p} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}\hat{q}/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n} \tag{7.10}$$

donde $\hat{q} = 1 - \hat{p}$ y como antes, el signo $(-)$ en la ecuación 7.10 corresponde al límite de confianza inferior y el signo $(+)$ al límite de confianza superior.

Esto se denomina a menudo como la *puntuación del intervalo de confianza* para p.

Si el tamaño n de la muestra es bastante grande, entonces $z^2/2n$ suele ser insignificante (pequeño) comparado con $\hat{p}$ y z^2/n es insignificante comparado con 1, partiendo de que $\tilde{p} \approx \hat{p}$. En este caso $z^2/4n^2$ también es despreciable comparado con pq/n (n^2 es un divisor mucho más grande que n); como resultado, el término dominante en la expresión $\pm$ es $z_{\alpha/2}$ es $\sqrt{\hat{p}\hat{q}/n}$ y el intervalo de puntuación es aproximadamente

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \tag{7.11}$$

Este último intervalo tiene la forma general $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$ de un amplio intervalo de la muestra sugerido en la última subsección. La aproximación del intervalo de confianza (7.11) es el que durante décadas ha aparecido en libros de texto de introducción a la estadística. Está claro que tiene una forma mucho más simple y más atractiva que la puntuación del intervalo de confianza. Así que, ¿por qué molestarse con este último?

Primero que todo, supóngase que se utiliza $z_{.025} = 1.96$ en la fórmula tradicional (7.11). Entonces, nuestro nivel de confianza *nominal* (el que creo que va a comprar utilizando este valor crítico z) es de aproximadamente 95%. Así que antes de seleccionar una muestra, la probabilidad de que el intervalo aleatorio incluya el valor real de p (es decir, la *probabilidad de cobertura*) debe ser de .95. Pero, como muestra la figura 7.6 para el caso $n = 100$, la probabilidad de cobertura real de este intervalo puede variar considerablemente de la probabilidad nominal de .95, en particular cuando p no está cerca de .5 (la gráfica de probabilidad de cobertura frente a p es muy irregular debido a que la distribución subyacente de probabilidad binomial es discreta y no continua). Esto es en general una deficiencia del intervalo tradicional, el nivel de confianza real puede ser bastante diferente del nivel nominal, incluso para tamaños de muestra razonablemente grandes. Investigaciones recientes han demostrado que el intervalo de la puntuación rectifica este comportamiento para prácticamente todos los tamaños de las muestras y los valores de p, su nivel de confianza real será bastante cercano al nivel nominal especificado por la elección de $z_{\alpha/2}$. Esto se debe en gran parte al hecho de que el intervalo de la puntuación se desplaza un poco hacia el .5 en comparación con los intervalos tradicionales. En particular, el punto medio $\tilde{p}$ del intervalo de la puntuación es siempre un poco más cercano a .5 que el punto medio $\hat{p}$ del intervalo tradicional. Esto es especialmente importante cuando p está cerca de 0 o 1.

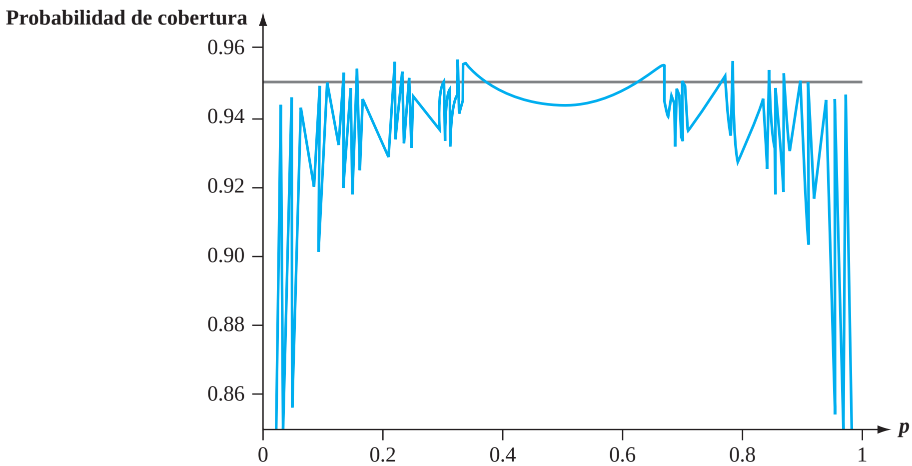


Figura 7.6 Probabilidad de cobertura real para el intervalo (7.11) para variaciones en los valores de p cuando $n = 100$

Además, el intervalo de la puntuación se puede utilizar con casi todos los tamaños de muestra y valores de los parámetros. No es necesario controlar las condiciones $n\hat{p} \geq 10$ y $n(1 - \hat{p}) \geq 10$ que se requerirían al emplear intervalos tradicionales. Así que en lugar de preguntar cuándo n es suficientemente grande para (7.11) para obtener una buena aproximación a (7.10), nuestra recomendación es que la puntuación del intervalo de confianza debe usarse *siempre*. El leve aburrimiento adicional de los cálculos se ve compensado por las propiedades deseables del intervalo.

Ejemplo 7.8 El artículo "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.*, 1997: 151–153) reportó que en $n = 48$ ensayos en un laboratorio particular, 16 dieron por resultado la ignición de un tipo particular de sustrato por un cigarrillo encendido. Sea p la proporción a largo plazo de todos los ensayos que producirían ignición. Una estimación puntual de p es $\hat{p} = 16/48 = .333$. Un intervalo de confianza para p con un nivel de confianza de aproximadamente 95% es

$$\frac{.333 + (1.96)^2/96}{1 + (1.96)^2/48} \pm (1.96)\frac{\sqrt{(.333)(.667)/48 + (1.96)^2/9216}}{1 + (1.96)^2/48}$$
$$= .345 \pm .129 = (.216, .474)$$

Este intervalo es bastante amplio ya que un tamaño de muestra de 48 no es tan grande al estimar una proporción.

El intervalo tradicional es

$$.333 \pm 1.96\sqrt{(.333)(.667)/48} = .333 \pm .133 = (.200, .466)$$

Estos dos intervalos concordarían mucho más si el tamaño de muestra fuera sustancialmente más grande. ■

Si se iguala el ancho del intervalo de confianza para p con el ancho preespecificado w se obtiene una ecuación cuadrática para el tamaño de muestra n necesario para dar un intervalo con un grado de precisión deseado. Si se suprime el subíndice en $z_{\alpha/2}$, la solución es

$$n = \frac{2z^2\hat{p}\hat{q} - z^2w^2 \pm \sqrt{4z^4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^2z^4}}{w^2} \tag{7.12}$$

Omitiendo los términos en el numerador que implican w^2 se obtiene

$$n \approx \frac{4z^2\hat{p}\hat{q}}{w^2}$$

Esta última expresión es lo que resulta de igualar el ancho del intervalo tradicional con w.

Estas fórmulas desafortunadamente implican la $\hat{p}$ desconocida. El método más conservador es aprovechar el hecho de que $\hat{p}\hat{q}[= \hat{p}(1 - \hat{p})]$ es un máximo cuando $\hat{p} = .5$. Por consiguiente, si se utiliza $\hat{p} = \hat{q} = .5$ en (7.12), el ancho será cuando mucho w haciendo caso omiso de qué valor de $\hat{p}$ resulte de la muestra. De manera alternativa, si el investigador cree de manera firme, basado en información previa, que $p \le p_0 \le .5$, en ese caso se utiliza p_0 en lugar de $\hat{p}$. Un comentario similar es válido cuando $p \ge p_0 \ge .5$.

Ejemplo 7.9 El ancho del intervalo de confianza de 95% en el ejemplo 7.8 es .258. El valor de n necesario para garantizar un ancho de .10 independientemente del valor de $\hat{p}$ es

$$n = \frac{2(1.96)^2(.25) - (1.96)^2(.01) \pm \sqrt{4(1.96)^4(.25)(.25 - .01) + (.01)(1.96)^4}}{.01} = 380.3$$

Por consiguiente se deberá utilizar un tamaño de muestra de 381. La expresión para n basada en el intervalo de confianza tradicional da un valor un poco más grande que 385. ■

Intervalos de confianza unilaterales (límites de confianza)

Los intervalos de confianza discutidos hasta ahora dan tanto un límite de confianza inferior *como* uno superior para el parámetro que se está estimando. En algunas circunstancias, es posible que un investigador desee sólo uno de estos dos tipos de límites. Por ejemplo, es posible que un psicólogo desee calcular un límite de confianza superior de 95% para el tiempo de reacción promedio verdadero a un estímulo particular o es posible que un ingeniero de seguridad desee sólo un límite de confianza inferior para la vida útil promedio real de componentes de un tipo. Como el área acumulativa bajo la curva normal estándar a la izquierda de 1.645 es de .95,

$$P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} < 1.645\right) \approx .95$$

Si se manipula la desigualdad entre paréntesis para aislar μ en un lado y se reemplazan las variables aleatorias con valores calculados se obtiene la desigualdad $\mu > \overline{x} - 1.645s/\sqrt{n}$; la expresión a la derecha es el límite de confianza inferior deseado. Comenzando con $P(-1.645 < Z) \approx .95$ y manipulando la desigualdad se obtiene el límite de confianza superior. Un argumento similar da un límite unilateral asociado con cualquier otro nivel de confianza.

PROPOSICIÓN

Un **límite de confianza superior muestral grande para μ** es

$$\mu < \overline{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$$

y un **límite de confianza inferior muestral grande para μ** es

$$\mu > \overline{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$$

Se obtiene **un límite de confianza unilateral para p** reemplazando $z_{\alpha/2}$ en lugar de z_α y $\pm$ en lugar de $+$ o $-$ en la fórmula para el intervalo de confianza (7.10) para p. En todos los casos, el nivel de confianza es aproximadamente de $100(1 - \alpha)\%$.

Ejemplo 7.10 La prueba de esfuerzo cortante es el procedimiento más aceptado para evaluar la calidad de una unión entre un material de reparación y su sustrato de concreto. El artículo "Testing the Bond Between Repair Materials and Concrete Substrate" (*ACI Materials J.*, 1996: 553–558) reportó que en una investigación particular, una muestra de 48 observaciones de resistencia al esfuerzo cortante dio una resistencia media muestral de 17.17 N/mm^2 y una desviación estándar muestral de 3.28 N/mm^2. Un límite de confianza inferior para la resistencia al esfuerzo cortante promedio verdadera μ con nivel de confianza de 95% es

$$17.17 - (1.645)\frac{(3.28)}{\sqrt{48}} = 17.17 - .78 = 16.39$$

Esto es, con un nivel de confianza de 95%, se puede decir que $\mu > 16.39$. ■

EJERCICIOS Sección 7.2 (12–27)

12. Una muestra aleatoria de 110 relámpagos en cierta región dieron por resultado una duración de eco de radar promedio muestral de .81 segundos y una desviación estándar muestral de .34 segundos ("Lightning Strikes to an Airplane in a Thunderstorm", *J. of Aircraft*, 1984: 607–611). Calcule un intervalo de confianza de 99% (bilateral) para la duración de eco promedio verdadera μ e interprete el intervalo resultante.

13. El artículo "Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products" (*Indoor Air*, 2006: 65–73) reportó que para una muestra de 50 cocinas con estufas de gas monitoreadas durante una semana, el nivel de CO_2 medio muestral (ppm) fue de 654.16 y la desviación estándar muestral fue de 164.43.

a. Calcule e interprete un intervalo de confianza de 95% (bilateral) para un nivel de CO_2 promedio verdadero en la población de todas las casas de la cual se seleccionó la muestra.

b. Suponga que el investigador había hecho una suposición preliminar de 175 para el valor de s antes de recopilar los datos. ¿Qué tamaño de muestra sería necesario para obtener un ancho de intervalo de 50 ppm para un nivel de confianza de 95%?

14. El artículo "Evaluating Tunnel Kiln Performance" (*Amer. Ceramic Soc. Bull.*, agosto de 1997: 59–63) reportó la siguiente información resumida sobre resistencias a la fractura (MPa) de $n = 169$ barras de cerámica horneadas en un horno particular: $\bar{x} = 89.10$, $s = 3.73$.

a. Calcule un intervalo de confianza (bilateral) para la resistencia a la fractura promedio verdadera utilizando un nivel de confianza de 95%. ¿Se podría decir que la resistencia a la fractura promedio verdadera fue estimada con precisión?

b. Suponga que los investigadores creyeron *a priori* que la desviación estándar de la población era aproximadamente de 4 MPa. Basado en esta suposición, ¿qué tan grande tendría que ser una muestra para estimar μ dentro de .5 MPa con 95% de confianza?

15. Determine el nivel de confianza de cada uno de los siguientes límites de confianza unilaterales muestrales grandes:

a. Límite superior: $\bar{x} + .84s/\sqrt{n}$

b. Límite inferior: $\bar{x} - 2.05s/\sqrt{n}$

c. Límite superior: $\bar{x} + .67s/\sqrt{n}$

16. El voltaje de ruptura de corriente alterna (AC) de un líquido aislante indica su rigidez dieléctrica. El artículo "Testing Practices for the AC Breakdown Voltage Testing of Insulation Liquids" (*IEEE Electrical Insulation Magazine*, 1995: 21−26) dio las observaciones muestrales adjuntas de voltaje de ruptura (kV) de un circuito particular, en ciertas condiciones.

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55
57	48	63	57	57	55	53	59	53	52	50	55	60	50	56	58

a. Construya un diagrama de caja de los datos y comente sobre las características interesantes.

b. Calcule e interprete un intervalo de confianza del 95% para el promedio real del voltaje de ruptura μ. ¿Parece que μ ha sido estimada con precisión? Explique.

c. Supongamos que el investigador cree que prácticamente todos los valores de voltaje de ruptura están entre 40 y 70. ¿Qué tamaño de la muestra sería conveniente para que el intervalo de confianza del 95% tenga una anchura de 2 kV (de modo que μ se estime dentro de 1 kV con 95% de confianza)?

17. El ejercicio 1.13 dio una muestra de observaciones de resistencia última a la tensión (kg/pulg^2). Use los datos de salida estadísticos descriptivos adjuntos de Minitab para calcular un límite de confianza inferior de 99% para la resistencia a la tensión última promedio verdadera e interprete el resultado.

N	Media	Mediana	TrMedia	DesvEst	ECMedia
153	135.39	135.40	135.41	4.59	0.37

Mínimo	Máximo	Q1	Q3
122.20	147.70	132.95	138.25

18. El artículo "Ultimate Load Capacities of Expansion Anchor Bolts" (*J. of Energy Engr.*, 1993: 139–158) reportó los siguientes datos resumidos sobre resistencia al esfuerzo cortante (kg/pulg^2) para una muestra de pernos de anclaje de 3/8 pulg: $n = 78$, $\bar{x} = 4.25$, $s = 1.30$. Calcule un límite de confianza inferior utilizando un nivel de confianza de 90% para una resistencia al esfuerzo cortante promedio verdadera.

19. El artículo "Limited Yield Estimation for Visual Defect Sources" (*IEEE Trans. on Semiconductor Manuf.*, 1997: 17–23) reportó que, en un estudio de un proceso de inspección de obleas particular, 356 troqueles fueron examinados por una sonda de inspección y 201 de éstos pasaron la prueba. Suponiendo un proceso estable, calcule un intervalo de confianza (bilateral) de 95% para la proporción de todos los troqueles que pasan la prueba.

20. La Prensa Asociada (9 de octubre de 2002) reportó que en una encuesta de 4722 jóvenes estadounidenses de 6 a 19 años de edad, 15% sufría de problemas serios de sobrepeso (un índice de masa corporal de por lo menos 30; este índice mide el peso con respecto a la estatura). Calcule e interprete un intervalo de confianza utilizando un nivel de confianza de 99% para la proporción de todos los jóvenes estadounidenses con un problema serio de sobrepeso.

21. En una muestra de 1000 consumidores seleccionados al azar que tuvieron la oportunidad de enviar un formulario de solicitud de reembolso después de comprar un producto, 250 de estas personas dijeron que nunca lo hicieron ("Rebates: Get What You Deserve", *Consumer Reports*, mayo de 2009: 7). Las razones citadas para su comportamiento incluyen demasiados pasos en el proceso, cantidad demasiado pequeña, vencimiento del plazo, el temor de ser puesto en una lista de correo, la pérdida de su recepción y las dudas acerca de recibir el dinero. Calcule un límite de confianza superior al nivel de confianza del 95% para la verdadera proporción de estos consumidores que nunca solicitaron un reembolso. Con base en este límite, ¿hay pruebas convincentes de que la verdadera proporción de estos consumidores es menor que 1/3? Explique su razonamiento.

22. La tecnología subyacente de reemplazos de cadera ha cambiado ya que estas operaciones se han vuelto más populares (más de 250,000 en Estados Unidos en el año 2008). A partir del año 2003, las caderas de cerámica de alta duración se comercializaban. Desafortunadamente, para muchos pacientes la mayor durabilidad ha sido compensada por un aumento en la incidencia de chirridos. La edición del 11 de mayo de 2008 del *New York Times* informó que en un estudio de 143 individuos que recibieron las caderas de cerámica entre los años 2003 y 2005, 10 de las caderas desarrollaron chirridos.

a. Calcule un límite de confianza inferior en el nivel de confianza del 95% para la verdadera proporción de las caderas que desarrollaron chirridos.

b. Interprete el nivel de confianza del 95% utilizado en el inciso (a).

23. El Pew Forum on Religion and Public Life reportó el 9 de diciembre del año 2009 que en una encuesta de 2003 adultos estadounidenses, 25% dijo que creía en la astrología.

a. Calcule e interprete un intervalo de confianza al nivel de confianza del 99% para la proporción de todos los adultos estadounidenses que creen en la astrología.

b. ¿Qué tamaño de muestra se requiere para que el ancho de un intervalo de confianza de 99% tenga un máximo de .05, independientemente del valor de $\hat{p}$?

24. Una muestra de 56 muestras de algodón produjo un porcentaje de alargamiento promedio muestral de 8.17 y una desviación estándar de 1.42 ("An Apparent Relation Between the Spiral Angle ϕ, the Percent Elongation E_1, and the Dimensions of the Cotton Fiber", *Textile Research J.*, 1978: 407–410). Calcule un intervalo de confianza de 95% muestral grande para el porcentaje de alargamiento promedio verdadero μ. ¿Qué suposiciones está haciendo sobre la distribución del porcentaje de alargamiento?

25. Una legisladora estatal desea encuestar a los residentes de su distrito para ver qué proporción del electorado está consciente de su posición sobre la utilización de fondos estatales para solventar abortos.

a. ¿Qué tamaño de muestra es necesario si el intervalo de confianza de 95% para p tiene que tener un ancho de cuando mucho .10 independientemente de p?

b. Si la legisladora está firmemente convencida de que por lo menos 2/3 del electorado conoce su posición, ¿qué tamaño de muestra recomendaría?

26. El superintendente de un gran distrito escolar, que una ocasión tomó un curso de probabilidad y estadística, cree que el número de maestros ausentes en cualquier día dado tiene una distribución de Poisson con parámetro μ. Use los datos adjuntos sobre ausencias durante 50 días para obtener un intervalo de confianza muestral grande para μ. [*Sugerencia*: la media y la varianza de una variable de Poisson son iguales a μ, por consiguiente

$$Z = \frac{\overline{X} - \mu}{\sqrt{\mu/n}}$$

tiene aproximadamente una distribución normal estándar. Ahora prosiga como en la deducción del intervalo para p haciendo una proposición de probabilidad (con probabilidad de $1 - \alpha$) y resolviendo las desigualdades resultantes para μ (véase el argumento exactamente después de (7.10).]

Número de ausencias	0	1	2	3	4	5	6	7	8	9	10
Frecuencia	1	4	8	10	8	7	5	3	2	1	1

27. Reconsidere el intervalo de confianza (7.10) para p y enfóquese en un nivel de confianza de 95%. Demuestre que los límites de confianza concuerdan bastante bien con los del intervalo tradicional (7.11) una vez que dos éxitos y dos fallas se anexaron a la muestra [es decir, (7.11) basado en $x + 2$ S en $n + 4$ ensayos]. [*Sugerencia*: $1.96 \approx 2$. *Nota*: Agresti y Coull demostraron que este ajuste del intervalo tradicional también tiene un nivel de confianza próximo al nivel nominal.]

7.3 Intervalos basados en una distribución de población normal

El intervalo de confianza para μ presentado en la sección 7.2 es válido siempre que n es grande. El intervalo resultante puede ser utilizado cualquiera que sea la naturaleza de la distribución de la población. El teorema del límite central no puede ser invocado, sin embargo, cuando n es pequeña. En este caso, una forma de proceder es hacer una suposición específica sobre la forma de la distribución de la población y luego deducir un intervalo de confianza adecuado a esa suposición. Por ejemplo, se podría desarrollar un intervalo de confianza para μ, cuando una distribución gamma describe la población, otro para el caso de una población de Weibull, y así sucesivamente. Estadísticos en realidad han realizado este programa para varias familias distribucionales diferentes. Como la distribución normal es más frecuentemente apropiada como modelo de población que cualquier otro tipo de distribución, la atención aquí se concentrará en un intervalo de confianza para esta situación.

SUPOSICIÓN

La población de interés es normal, de modo que $X_1, \ldots, X_n$ constituyen una muestra aleatoria tomada de una distribución normal con μ y σ desconocidas.

El resultado clave que subyace en el intervalo de la sección 7.2 fue que con n grande, la variable aleatoria $Z = (\overline{X} - \mu)/(S/\sqrt{n})$ tiene aproximadamente una distribución normal estándar. Cuando n es pequeño, no es probable que S se aproxime a σ, de modo que la variabilidad de la distribución de Z surge de la aleatoriedad tanto en el numerador como en el denominador. Esto implica que la distribución de probabilidad de $(\overline{X} - \mu)/(S/\sqrt{n})$ se dispersará más que la distribución normal estándar. El resultado en el cual están basadas las inferencias introduce una nueva familia de distribuciones de probabilidad llamada *distribuciones t*.

TEOREMA

Cuando $\overline{X}$ es la media de una muestra aleatoria de tamaño n tomada de una distribución normal con media μ, la variable aleatoria

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{7.13}$$

tiene una distribución de probabilidad llamada distribución t con $n - 1$ grados de libertad (gl).

Propiedades de distribuciones *t*

Antes de aplicar este teorema, se impone una discusión de las propiedades de distribuciones t. Aunque la variable de interés sigue siendo $(\overline{X} - \mu)/(S/\sqrt{n})$, ahora se denota por T para recalcar que no tiene una distribución normal estándar cuando n es pequeña. Recuérdese que una distribución normal está regida por dos parámetros, cada elección diferente de μ en combinación con σ resulta en una distribución normal particular. Cualquier distribución t particular resulta de especificar el valor de sólo un parámetro, llamado **número de grados de libertad**, abreviado como gl. Este parámetro se denota con la letra griega ν. Posibles valores de ν son los enteros positivos 1, 2, 3, . . . Así que hay una distribución t con un gl, otra con 2 gl, otra con 3 gl, y así sucesivamente.

Para cualquier valor fijo del parámetro ν, la función de densidad que especifica la curva t asociada es incluso más complicada que la función de densidad normal. Afortunadamente, sólo hay que ocuparse de algunas de las más importantes características de estas curvas.

Propiedades de distribuciones *t*

Sea t_ν, que denota la distribución t con ν gl.

1. Cada curva t_ν tiene forma de campana y su centro en 0.
2. Cada curva t_ν está más esparcida que la curva (z) normal estándar.
3. Conforme ν se incrementa, la dispersión de la curva t_ν correspondiente disminuye.
4. A medida que $\nu \to \infty$, la secuencia de curvas t_ν tiende a la curva normal estándar (así que la curva z a menudo se llama curva t con grados de libertad $= \infty$).

La figura 7.7 ilustra varias de estas propiedades para valores seleccionados de ν.

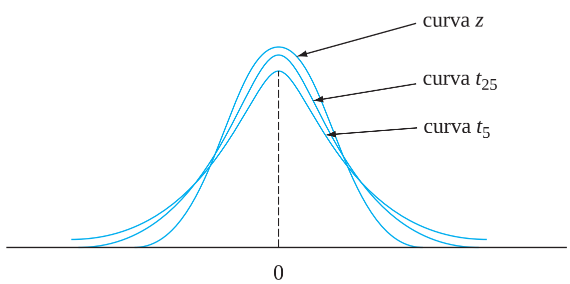


Figura 7.7 Curvas t_ν y z

El número de grados de libertad para T en (7.13) es $n - 1$ porque, aunque S está basada en las n desviaciones $X_1 - \overline{X}, \ldots, X_n - \overline{X}$, $\Sigma(X_i - \overline{X}) = 0$ implica que sólo

$n - 1$ de éstas están “libremente determinadas”. El número de grados de libertad para una variable t es el número de desviaciones libremente determinadas en las cuales está basada la desviación estándar estimada en el denominador de T.

El uso de la distribución t al hacer inferencias requiere notación para capturar áreas de cola de la curva t análogas a z_α de la curva z. Se podría pensar que t_α haría el truco. Sin embargo, el valor deseado depende no sólo del área de la cola capturada, sino también de gl.

NOTACIÓN

Sea $t_{\alpha,\nu}$ = el número sobre el eje de medición con el cual el área bajo la curva t con ν grados de libertad a la derecha de $t_{\alpha,\nu}$ es α; $t_{\alpha,\nu}$ se llama **valor crítico *t*.**

Por ejemplo, $t_{.05,6}$ es el valor crítico t que captura un área de cola superior de .05 bajo la curva t con 6 gl. La notación general se ilustra en la figura 7.8. Debido a que las curvas t son simétricas alrededor de cero, $-t_{\alpha,\nu}$ captura el área α de la cola inferior. La tabla A.5 del apéndice da $t_{\alpha,\nu}$ para valores seleccionados de α y ν. Esta tabla también aparece al final del libro. Las columnas de la tabla corresponden a diferentes valores de α. Para obtener $t_{.05,15}$, vaya a la columna $\alpha = .05$, mire hacia abajo al renglón $\nu = 15$, y lea $t_{.05,15} = 1.753$. Del mismo modo, $t_{.05,22} = 1.717$ (columna .05, renglón $\nu = 22$) y $t_{.01,22} = 2.508$.

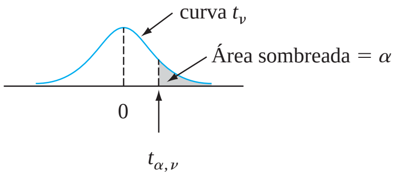


Figura 7.8 Ilustración de un valor crítico *t*

Los valores de $t_{\alpha,\nu}$ exhiben un comportamiento regular al recorrer una fila o al descender por una columna. Con ν fijo, $t_{\alpha,\nu}$ se incrementa a medida que α disminuye, puesto que hay que moverse más a la derecha de cero para capturar el área α en la cola. Con α fija, a medida que ν se incrementa (es decir, cuando se recorre hacia abajo cualquier columna particular de la tabla t) el valor de $t_{\alpha,\nu}$ disminuye. Esto es porque un valor más grande de ν implica una distribución t con dispersión más pequeña, de modo que no es necesario ir más lejos de cero para capturar el área de cola α. Además, $t_{\alpha,\nu}$ disminuye más lentamente a medida que ν se incrementa. Por consiguiente, los valores que aparecen en la tabla se muestran en incrementos de 2 entre 30 grados de libertad y 40 grados de libertad y luego saltan a $\nu = 50, 60, 120$ y por último ∞. Como t_∞ es la curva normal estándar, los valores z_α conocidos aparecen en la última fila de la tabla. La regla empírica sugería con anterioridad que el uso del intervalo de confianza muestral grande (si $n > 40$) proviene de la igualdad aproximada de las distribuciones normales estándar y t para $\nu \geq 40$.

Intervalo de confianza *t* para una muestra

La variable estandarizada T tiene una distribución t con $n - 1$ grados de libertad y el área bajo la curva de densidad t correspondiente entre $-t_{\alpha/2,n-1}$ y $t_{\alpha/2,n-1}$ es $1 - \alpha$ (el área $\alpha/2$ queda en cada cola), por consiguiente

$$P(-t_{\alpha/2,\,n-1} < T < t_{\alpha/2,\,n-1}) = 1 - \alpha \tag{7.14}$$

La expresión (7.14) difiere de las expresiones que aparecen en secciones previas en que T y $t_{\alpha/2,n-1}$ se utilizan en lugar de Z y $z_{\alpha/2}$, aunque pueden ser manipuladas de la misma manera para obtener un intervalo de confianza para μ.

PROPOSICIÓN

Sean $\bar{x}$ y s la media y la desviación estándar muestrales calculadas con los resultados de una muestra aleatoria tomada de una población normal con media μ. Entonces un **intervalo de confianza de 100(1 − α)% para μ** es

$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right) \tag{7.15}$$

o, más compactamente, $\bar{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$.

Un **límite de confianza superior para μ** es

$$\bar{x} + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$

y reemplazando $+$ por $-$ en la última expresión se obtiene un **límite de confianza inferior para μ,** ambos con nivel de confianza de $100(1 - \alpha)\%$.

Ejemplo 7.11 A pesar de que los mercados tradicionales de madera de ocozol han disminuido, las maderas sólidas de gran sección utilizadas tradicionalmente para la construcción de puentes y duelas se han vuelto cada vez más escasas. El artículo "Development of Novel Industrial Laminated Planks from Sweetgum Lumber" (*J. of Bridge Engr.*, 2008: 64–66) describe la fabricación y el ensayo de vigas compuestas, concebida para agregar valor a la madera de ocozol de bajo grado. Aquí hay datos sobre el módulo de ruptura (psi; el artículo contenía datos resumidos, expresados en MPa):

6807.99	7637.06	6663.28	6165.03	6991.41	6992.23
6981.46	7569.75	7437.88	6872.39	7663.18	6032.28
6906.04	6617.17	6984.12	7093.71	7659.50	7378.61
7295.54	6702.76	7440.17	8053.26	8284.75	7347.95
7422.69	7886.87	6316.67	7713.65	7503.33	7674.99

La figura 7.9 muestra un diagrama de probabilidad normal obtenido con el software R. La derechura del patrón en el diagrama apoya fuertemente la suposición de que la distribución de la población del módulo de ruptura es por lo menos aproximadamente normal.

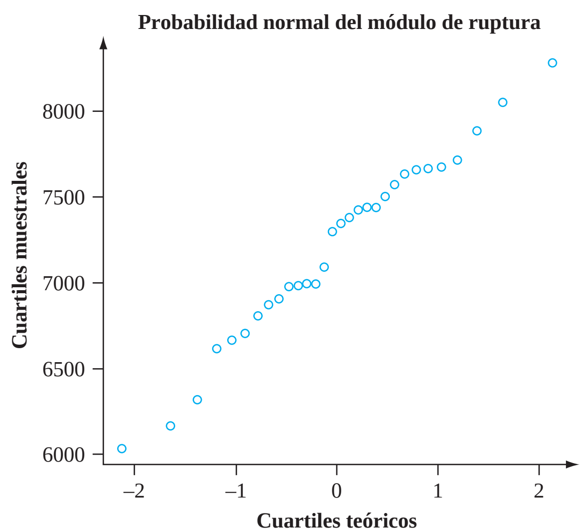


Figura 7.9 Diagrama de probabilidad normal de los datos del módulo de ruptura

La media muestral y la desviación estándar de la muestra son 7203.191 y 543.5400, respectivamente (para abatir cualquier realización de cálculos a mano, la carga computacional se alivia un poco al restar 6000 de cada valor de x para obtener $y_i = x_i - 6000$, entonces $\sum y_i = 36{,}095.72$ y $\sum y_i^2 = 51{,}997{,}668.77$, de la cual $\bar{y} = 1203.191$ y $s_y = s_x$ tal como se indica).

Ahora se calcula un intervalo de confianza para el promedio real del módulo de ruptura con un nivel de confianza de 95%. El intervalo de confianza se basa en $n - 1 = 29$ grados de libertad, por lo que el valor de t crítico necesario es $t_{.025,29} = 2.045$. La estimación por intervalo es ahora

$$\begin{aligned} \bar{x} \pm t_{.025,29} \cdot \frac{s}{\sqrt{n}} &= 7203.191 \pm (2.045) \cdot \frac{543.5400}{\sqrt{30}} \\ &= 7203.191 \pm 202.938 = (7000.253, 7406.129) \end{aligned}$$

Se estima que $7000.253 < \mu < 7406.129$ con un 95% de confianza. Si se utiliza la misma fórmula en muestra tras muestra, en el largo plazo el 95% de los intervalos calculados contendrán a μ. Dado que el valor de μ no está disponible, no sabemos si el intervalo calculado es uno de los "buenos" del 95% o el "malo" del 5%. Incluso con el tamaño de la muestra moderadamente grande, el intervalo es bastante amplio. Esto es una consecuencia de la cantidad sustancial de variabilidad de la muestra en los valores del módulo de ruptura.

Un límite de confianza inferior al 95% resultaría de conservar únicamente el límite de confianza inferior (el que tiene el signo $(-)$) y reemplazar 2.045 con $t_{.05,29} = 1.699$. ■

Por desgracia, no es fácil seleccionar n para controlar el ancho del intervalo t. Esto es porque el ancho implica la s desconocida (antes de recopilar los datos) y porque n ingresa no sólo a través de $1/\sqrt{n}$ sino también a través de $t_{\alpha/2,n-1}$. Por consiguiente, se puede obtener una n apropiada sólo mediante ensayo y error.

En el capítulo 15 se discutirá un intervalo de confianza de muestra pequeña para μ que es válido siempre que la distribución de la población sea simétrica, una suposición más débil que la de normalidad. No obstante, cuando la distribución de la población es normal, el intervalo t tiende a acortarse más de lo que lo haría *cualquier* otro intervalo con el mismo nivel de confianza.

Un intervalo de predicción para un solo valor futuro

En muchas aplicaciones, el objetivo es *predecir* un solo valor de una variable que tiene que ser observada en un tiempo futuro, en lugar de *estimar* el valor medio de dicha variable.

Ejemplo 7.12 Considere la siguiente muestra de contenido de grasa (en porcentaje) de $n = 10$ perros calientes seleccionados al azar ("Sensory and Mechanical Assessment of the Quality of Frankfurters", *J. of Texture Studies*, 1990: 395–409):

25.2	21.3	22.8	17.0	29.8	21.0	25.5	16.0	20.9	19.5

Suponiendo que estas observaciones se seleccionaron de una distribución de población normal, un intervalo de confianza de 95% para (estimación del intervalo de) el contenido de grasa medio de la población es

$$\begin{aligned} \bar{x} \pm t_{.025,9} \cdot \frac{s}{\sqrt{n}} &= 21.90 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = 21.90 \pm 2.96 \\ &= (18.94, 24.86) \end{aligned}$$

Suponga, sin embargo, que se va a comer un solo perro caliente de este tipo y desea *predecir* el contenido de grasa resultante. Una predicción *puntual*, análoga a una estimación *puntual*, es simplemente $\bar{x} = 21.90$. Esta predicción desafortunadamente no da información sobre confiabilidad o precisión. ■

El escenario general es como sigue: se dispone de una muestra aleatoria X_1, $X_2, \ldots, X_n$ tomada de una distribución de población normal y se desea predecir el valor de X_{n+1}, una sola observación futura (por ejemplo, la vida útil de un foco sencillo que se compra o la eficiencia de combustible de un automóvil simple que es rentado). Un predictor puntual es $\overline{X}$ y el error de predicción resultante es $\overline{X} - X_{n+1}$. El valor esperado del error de predicción es

$$E(\overline{X} - X_{n+1}) = E(\overline{X}) - E(X_{n+1}) = \mu - \mu = 0$$

Como X_{n+1}, es independiente de $X_1, \ldots, X_n$, es independiente de $\overline{X}$, así que la varianza del error de predicción es

$$V(\overline{X} - X_{n+1}) = V(\overline{X}) + V(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

El error de predicción es una combinación lineal de variables aleatorias independientes normalmente distribuidas, así que también está normalmente distribuido. Por consiguiente

$$Z = \frac{(\overline{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \dfrac{1}{n}\right)}} = \frac{\overline{X} - X_{n+1}}{\sqrt{\sigma^2\left(1 + \dfrac{1}{n}\right)}}$$

tiene una distribución normal estándar. Se puede demostrar que si se reemplaza σ con la desviación estándar muestral S (de $X_1, \ldots, X_n$) se obtiene

$$T = \frac{\overline{X} - X_{n+1}}{S\sqrt{1 + \dfrac{1}{n}}} \sim t \text{ distribución } t \text{ con } n - 1 \text{ grados de libertad}$$

Si se manipula esta variable T como se manipuló $T = (\overline{X} - \mu)/(S/\sqrt{n})$ en el desarrollo de un intervalo de confianza se obtiene el siguiente resultado.

PROPOSICIÓN

Un **intervalo de predicción** (IP) para una sola observación que tiene que ser seleccionado de una distribución de población normal es

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}} \qquad (7.16)$$

El *nivel de predicción* es $100(1 - \alpha)\%$. Una predicción del límite inferior resulta de la sustitución de $t_{\alpha/2}$ por t_α y desechar la parte + de (7.16), una modificación similar da una predicción del límite superior.

La interpretación de un nivel de predicción de 95% es similar a la de un nivel de confianza de 95%; si se calcula el intervalo (7.16) para muestra tras muestra, a la larga el 95% de estos intervalos incluirán los valores futuros correspondientes de X.

Ejemplo 7.13 (Continuación del ejemplo 7.12)

Con $n = 10$, $\bar{x} = 21.90$, $s = 4.134$ y $t_{.025,9} = 2.262$, un intervalo de predicción de 95% para el contenido de grasa de un solo perro caliente es

$$\begin{aligned} 21.90 \pm (2.262)(4.134)\sqrt{1 + \frac{1}{10}} &= 21.90 \pm 9.81 \\ &= (12.09, 31.71) \end{aligned}$$

El intervalo es bastante ancho, lo que indica una incertidumbre sustancial en cuanto al contenido de grasa. Obsérvese que el ancho del intervalo de predicción es más de tres veces el del intervalo de confianza. ■

El error de predicción es $\overline{X} - X_{n+1}$, la diferencia entre dos variables aleatorias, en tanto que el error de estimación es $\overline{X} - \mu$, la diferencia entre una variable aleatoria y un valor fijo (aunque desconocido). El intervalo de predicción es más ancho que el intervalo de confianza porque hay más variabilidad en el error de predicción (debido a X_{n+1}) que en el error de estimación. De hecho, a medida que n se hace arbitrariamente grande, el intervalo de confianza se contrae a un solo valor μ y el intervalo de predicción tiende a $\mu \pm z_{\alpha/2} \cdot \sigma$. Existe incertidumbre con respecto a un solo valor X incluso cuando no hay necesidad de estimarlo.

Intervalos de tolerancia

Considérese una población de automóviles de cierto tipo y supóngase que en condiciones específicas, la eficiencia de combustible (mpg) tiene una distribución normal con $\mu = 30$ y $\sigma = 2$. Entonces como el intervalo de -1.645 a 1.645 captura 90% del área bajo la curva z, 90% de todos estos automóviles tendrán valores de eficiencia de combustible entre $\mu - 1.645\sigma = 26.71$ y $\mu + 1.645\sigma = 33.29$. Pero, ¿qué sucederá si los valores de μ y σ no son conocidos? Se puede tomar una muestra de tamaño n, determinar las eficiencias de combustible, $\bar{x}$ y s, y formar el intervalo cuyo límite inferior es $\bar{x} - 1.645s$ y cuyo límite superior es $\bar{x} + 1.645s$. Sin embargo, debido a la variabilidad de muestreo en las estimaciones de μ y σ, existe una buena probabilidad de que el intervalo resultante incluya menos de 90% de los valores de la población. Intuitivamente, para tener *a priori* una probabilidad de que 95% del intervalo resultante incluya por lo menos 90% de los valores de la población, cuando $\bar{x}$ y s se utilizan en lugar de μ y σ, también se deberá reemplazar 1.645 con un número más grande. Por ejemplo, cuando $n = 20$, el valor 2.310 es tal que se puede estar 95% confiado en que el intervalo $\bar{x} \pm 2.310s$ incluirá por lo menos 90% de los valores de eficiencia de combustible en la población.

Sea k un número entre 0 y 100. Un **intervalo de tolerancia** para capturar por lo menos el k% de los valores en una distribución de población normal con nivel de confianza de 95% tiene la forma

$$\bar{x} \pm (\text{valor crítico de tolerancia}) \cdot s$$

En la tabla A.6 del apéndice aparecen valores críticos de tolerancia para $k = 90$, 95 y 99 en combinación con varios tamaños de muestra. Esta tabla también incluye valores críticos para un nivel de confianza de 99% (estos valores son más grandes que los valores correspondientes al 95%). Si se reemplaza $\pm$ con $+$ se obtiene un límite de tolerancia superior, y si se utiliza $-$ en lugar de $\pm$ se obtiene un límite de tolerancia inferior. En la tabla A.6 también aparecen valores críticos para obtener estos límites unilaterales.

Ejemplo 7.14 Como parte de un proyecto más amplio para estudiar el comportamiento de los paneles de corteza comprimida, un componente estructural que se utiliza ampliamente en América del Norte, el artículo "Time-Dependent Bending Properties of Lumber" (*J. of Testing and Eval.*, 1996: 187–193) informó sobre varias propiedades mecánicas de muestras de madera de pino escocés. Considere las siguientes observaciones sobre el módulo de elasticidad (MPa) obtenido un minuto después de la carga en una cierta configuración:

10,490	16,620	17,300	15,480	12,970	17,260	13,400	13,900
13,630	13,260	14,370	11,700	15,470	17,840	14,070	14,760

Hay un patrón lineal pronunciado en el gráfico de probabilidad normal de los datos. Un resumen de las cantidades importantes es $n = 16$, $\bar{x} = 14{,}532.5$, $s = 2055.67$. Para un nivel de confianza del 95%, un intervalo de tolerancia bilateral para la captura de al menos el 95% de los valores de los módulos elasticidad de las muestras de madera en la población de la muestra, utiliza el valor de tolerancia crítica de 2.903. El intervalo resultante es

$$14{,}532.5 \pm (2.903)(2055.67) = 14{,}532.5 \pm 5967.6 = (8{,}564.9,\ 20{,}500.1)$$

Se puede estar totalmente confiado de que por lo menos 95% de todos los especímenes de madera tienen valores de módulo de elasticidad de entre 8564.9 y 20,500.1.

El intervalo de confianza de 95% para μ fue (13,437.3, 15,627.7) y el intervalo de predicción de 95% para el módulo de elasticidad de un solo espécimen de madera es (10,017.0, 19,048.0). Tanto el intervalo de predicción como el intervalo de tolerancia son sustancialmente más anchos que el intervalo de confianza. ■

Intervalos basados en distribuciones de población no normales

El intervalo de confianza t para una muestra de μ es robusto en cuanto a alejamientos pequeños o incluso moderados de la normalidad a menos que n sea bastante pequeño. Con esto se quiere decir que si se utiliza un valor crítico para confianza de 95%, por ejemplo, al calcular el intervalo, el nivel de confianza real se aproximará de manera razonable al nivel nominal de 95%. Si, sin embargo, n es pequeño y la distribución de la población es altamente no normal, entonces el nivel de confianza real puede ser diferente en forma considerable del que se utiliza cuando se obtiene un valor crítico particular de la tabla t. Ciertamente, ¡sería penoso creer que el nivel de confianza es de más o menos 95% cuando en realidad es como de 88%! Se ha visto que la técnica bootstrap, introducida en la sección 7.1 es bastante exitosa al estimar parámetros en una amplia variedad de situaciones no normales.

En contraste con el intervalo de confianza, la validez de los intervalos de predicción y tolerancia descritos en esta sección está estrechamente vinculada a la suposición de normalidad. Estos últimos intervalos no deberán ser utilizados sin evidencia apremiante de normalidad. La excelente referencia *Statistical Intervals*, citada en la bibliografía al final de este capítulo, discute procedimientos alternativos de esta clase en varias otras situaciones.

EJERCICIOS Sección 7.3 (28–41)

28. Determine los valores de las siguientes cantidades:
a. $t_{.1,15}$ b. $t_{.05,15}$ c. $t_{.05,25}$ d. $t_{.05,40}$ e. $t_{.005,40}$

29. Determine el valor o valores crítico(os) t que capturará el área deseada de la curva t en cada uno de los siguientes casos:
a. Área central = .95, gl = 10
b. Área central = .95, gl = 20
c. Área central = .99, gl = 20
d. Área central = .99, gl = 50
e. Área de cola superior = .01, gl = 25
f. Área de cola inferior = .025, gl = 5

30. Determine el valor crítico t de un intervalo de confianza bilateral en cada una de las siguientes situaciones:
a. Nivel de confianza = 95%, gl = 10
b. Nivel de confianza = 95%, gl = 15
c. Nivel de confianza = 99%, gl = 15
d. Nivel de confianza = 99%, n = 5
e. Nivel de confianza = 98%, gl = 24
f. Nivel de confianza = 99%, n = 38

31. Determine el valor crítico t para un límite de confianza inferior o superior en cada una de las situaciones descritas en el ejercicio 30.

32. De acuerdo con el artículo "Fatigue Testing of Condoms" (*Polymer Testing*, 2009: 567−571), "las pruebas que se utilizan actualmente para los condones son sustitutos de los desafíos que enfrentan en uso", incluyendo una prueba de hoyos, una prueba de inflación, una prueba de sello del paquete y las pruebas de las dimensiones y la calidad del lubricante (¡todo el territorio fértil para el uso de la metodología estadística!). Los investigadores desarrollaron una nueva prueba que agrega tensión cíclica a un nivel muy por debajo de la rotura y determina el número de ciclos hasta llegar a la ruptura. Una muestra de 20 condones de un tipo particular, resultó en una media muestral de 1584 y una desviación estándar muestral de 607. Calcule e interprete un intervalo de confianza al nivel de confianza del 99% para el verdadero número promedio de ciclos de ruptura. [*Nota*: el artículo presenta los resultados de las pruebas de hipótesis basadas en la distribución t, la validez de éstas depende de suponer la distribución normal de la población.]

33. El artículo "Measuring and Understanding the Aging of Kraft Insulating Paper in Power Transformers" (*IEEE Electrical Insul. Mag.*, 1996: 28–34) contiene las siguientes observaciones del grado de polimerización de especímenes de papel para los cuales la concentración de tiempos de viscosidad cayeron en un rango medio:

418	421	421	422	425	427	431
434	437	439	446	447	448	453
454	463	465				

a. Construya una gráfica de caja de los datos y comente sobre cualquier característica interesante.
b. ¿Es plausible que las observaciones muestrales dadas fueran seleccionadas de una distribución normal?
c. Calcule un intervalo de confianza de 95% bilateral para un grado de polimerización promedio verdadero (como lo hicieron los autores del artículo). ¿Sugiere este intervalo que 440 es un valor factible del grado de polimerización promedio verdadero? ¿Qué hay en cuanto a 450?

34. Una muestra de 14 especímenes de junta de un tipo particular produjo un esfuerzo límite proporcional medio muestral de 8.48 MPa y una desviación estándar muestral de .79 MPa ("Characterization of Bearing Strength Factors in Pegged Timber Connections", *J. of Structural Engr*., 1997: 326–332).
a. Calcule e interprete un límite de confianza inferior de 95% para el esfuerzo límite proporcional promedio verdadero de todas las juntas. ¿Qué suposiciones, si hay alguna, hizo sobre la distribución del esfuerzo límite proporcional?
b. Calcule e interprete un límite de predicción inferior de 95% para el esfuerzo límite proporcional de una sola unión de este tipo.

35. Para corregir deformidades nasales congénitas se utiliza rinoplastia de aumento mediante implante de silicón. El éxito del procedimiento depende de varias propiedades biomecánicas del periostio y fascia nasales humanas. El artículo "Biomechanics in Augmentation Rhinoplasty" (*J. of Med. Engr. and Tech*., 2005: 14-17) reportó que para una muestra de 15 adultos (recién fallecidos), la deformación de falla media (en porcentaje) fue de 25.0 y la desviación estándar fue de 3.5.
a. Suponiendo una distribución normal de la deformación de falla, estime la deformación promedio verdadera en una forma que transmita información acerca de precisión y confiabilidad.
b. Pronostique la deformación para un solo adulto en una forma que transmita información sobre precisión y confiabilidad. ¿Cómo se compara la predicción con la estimación calculada en el inciso (a)?

36. Las $n = 26$ observaciones de tiempo de escape dadas en el ejercicio 36 del capítulo 1 dan una media y desviación estándar muestrales de 370.69 y 24.36, respectivamente.
a. Calcule un límite de confianza superior para el tiempo de escape medio de la población utilizando un nivel de confianza de 95%.
b. Calcule un límite de predicción superior para el tiempo de escape de un solo trabajador adicional utilizando un nivel de predicción de 95%. ¿Cómo se compara este límite con el límite de confianza del inciso (a)?
c. Suponga que se escogerán dos trabajadores más para participar en el ejercicio de escape simulado. Denote sus tiempos de escape por X_{27} y X_{28} y sea $\overline{X}_{\text{nuevo}}$ el promedio de estos dos valores. Modifique la fórmula para un intervalo de predicción con un solo valor de x para obtener un intervalo de predicción para $\overline{X}_{\text{nuevo}}$ y calcule un intervalo bilateral de 95% basado en los datos de escape dados.

37. Un estudio de la capacidad de individuos de caminar en línea recta ("Can We Really Walk Straight?" *Amer. J. of Physical Anthro*., 1992: 19-27) reportó los datos adjuntos sobre cadencia (pasos por segundo) con una muestra de $n = 20$ hombres saludables seleccionados al azar.

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

Un diagrama de probabilidad normal apoya de manera sustancial la suposición de que la distribución de la población de cadencia es aproximadamente normal. A continuación se da un resumen descriptivo de los datos obtenidos con Minitab:

```
Cadencia N   Media   Mediana  TrMedia  DesvEst  ECMedia
variable 20  0.9255  0.9300   0.9261   0.0809   0.0181

Cadencia       Mín      Máx      Q1       Q3
variable     0.7800   1.0600   0.8525   0.9600
```

a. Calcule e interprete un intervalo de confianza de 95% para la cadencia media de la población.
b. Calcule e interprete un intervalo de predicción de 95% para la cadencia de un solo individuo seleccionado al azar de esta población.
c. Calcule un intervalo que incluya por lo menos 99% de las cadencias en la distribución de la población utilizando un nivel de confianza de 95%.

38. Se seleccionó una muestra de 25 piezas de laminado utilizado en la fabricación de tarjetas de circuito y se determinó la cantidad de pandeo (pulg) en condiciones particulares para cada pieza y el resultado fue un pandeo medio muestral de .0635 y una desviación estándar muestral de .0065.
a. Calcule una predicción de la cantidad de pandeo de una sola pieza de laminado en una manera que proporcione información sobre precisión y confiabilidad.
b. Calcule un intervalo con el cual pueda tener un alto grado de confianza de que por lo menos 95% de todas las piezas de laminado produzcan cantidades de pandeo que estén entre los dos límites del intervalo.

39. El ejercicio 72 del capítulo 1 dio las siguientes observaciones en una medición de afinidad de receptor (volumen de distribución ajustado) con una muestra de 13 individuos sanos: 23, 39, 40, 41, 43, 47, 51, 58, 63, 66, 67, 69, 72.
a. ¿Es plausible que la distribución de la población de la cual se seleccionó esta muestra sea normal?
b. Calcule un intervalo con el cual pueda estar 95% confiado de que por lo menos 95% de todos los individuos saludables en la población tienen volúmenes de distribución ajustados que quedan entre los límites del intervalo.
c. Pronostique el volumen de distribución ajustado de un solo individuo saludable calculando un intervalo de predicción de 95%. ¿Cómo se compara el ancho de este intervalo con el ancho del intervalo calculado en el inciso (b)?

40. El ejercicio 13 del capítulo 1 presentó una muestra de $n = 153$ observaciones de resistencia última a la tensión y el ejercicio 17 de la sección previa dio cantidades resumidas y solicitó un intervalo de confianza muestral grande. Como el tamaño de la muestra es grande, no se requieren suposiciones sobre la distribución de la población en cuanto a la validez del intervalo de confianza.

a. ¿Se requiere alguna suposición sobre la distribución de la resistencia a la tensión antes de calcular un límite de predicción inferior para la resistencia a la tensión del nuevo espécimen seleccionado por medio del método descrito en esta sección? Explique.

b. Use un paquete de software estadístico para investigar la factibilidad de una distribución de población normal.

c. Calcule un límite de predicción inferior con un nivel de predicción de 95% para la resistencia última a la tensión del siguiente espécimen seleccionado.

41. Una tabulación más extensa de valores críticos t que la que aparece en este libro muestra que para la distribución t con 20 grados de libertad, las áreas a la derecha de los valores .687, .860 y 1.064 son .25, .20 y .15, respectivamente. ¿Cuál es el nivel de confianza de cada uno de los siguientes tres intervalos de confianza para la media μ de una distribución de población normal? ¿Cuál de los tres intervalos recomendaría utilizar y por qué?

a. $(\bar{x} - .687s/\sqrt{21}, \bar{x} + 1.725s/\sqrt{21})$
b. $(\bar{x} - .860s/\sqrt{21}, \bar{x} + 1.325s/\sqrt{21})$
c. $(\bar{x} - 1.064s/\sqrt{21}, \bar{x} + 1.064s/\sqrt{21})$

7.4 Intervalos de confianza para la varianza y desviación estándar de una población normal

Aun cuando las inferencias en lo que se refiere a la varianza σ^2 o a la desviación estándar de una población en general son de menor interés que aquellas con respecto a una media o proporción, hay ocasiones en que se requieren tales procedimientos. En el caso de una distribución de población normal, las inferencias están basadas en el siguiente resultado por lo que se refiere a la varianza muestral S^2.

TEOREMA

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con parámetros μ y σ^2. Entonces la variable aleatoria

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$$

tiene una distribución de probabilidad ji cuadrada (χ^2) con $n - 1$ grados de libertad.

Como se discutió en las secciones 4.4 y 7.1, la distribución ji cuadrada es una distribución de probabilidad continua con un solo parámetro ν, llamado número de grados de libertad, con posibles valores de 1, 2, 3, . . . Las gráficas de varias funciones de densidad de probabilidad χ^2 se ilustran en la figura 7.10. Cada función de densidad de probabilidad $f(x; \nu)$ es positiva sólo para $x > 0$, y cada una tiene asimetría positiva (una larga cola superior), aunque la distribución se mueve hacia la derecha y se vuelve más simétrica a medida que se incrementa ν. Para especificar procedimientos inferenciales que utilizan la distribución ji cuadrada, se requiere una notación análoga a aquella para un valor t crítico $t_{\alpha,\nu}$.

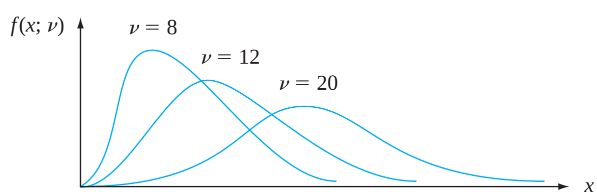


Figura 7.10 Gráficas de funciones de densidad ji cuadrada

NOTACIÓN

Sea $\chi^2_{\alpha,\nu}$, llamado **valor crítico ji cuadrada,** el número sobre el eje horizontal de modo que α del área bajo la curva ji cuadrada con ν grados de libertad quede a la derecha de $\chi^2_{\alpha,\nu}$.

La simetría de las distribuciones t hizo que fuera necesario tabular sólo valores críticos t de cola superior ($t_{\alpha,\nu}$ con valores pequeños de α). La distribución ji cuadrada no es simétrica, por lo que la tabla A.7 del apéndice contiene valores de $\chi^2_{\alpha,\nu}$ tanto para α cerca de 0 como cerca de 1, como se ilustra en la figura 7.11(b). Por ejemplo, $\chi^2_{.025,14} = 26.119$, y $\chi^2_{.95,20}$ (el 5º percentil) $= 10.851$.

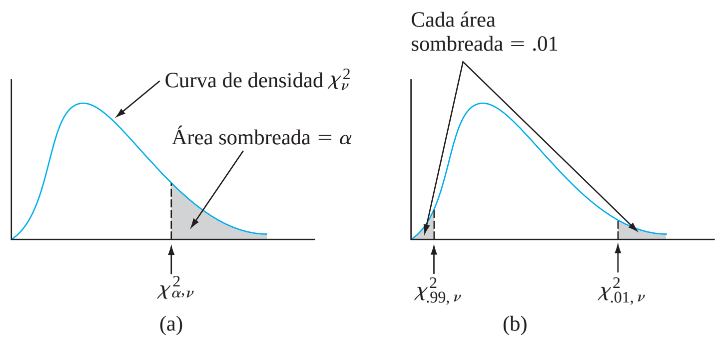


Figura 7.11 Notación $\chi^2_{\alpha,\nu}$ ilustrada

La variable aleatoria $(n-1)S^2/\sigma^2$ satisface las dos propiedades en las cuales está basado el método general para obtener un intervalo de confianza: es una función del parámetro de interés σ^2, no obstante su distribución de probabilidad (ji cuadrada) no depende de este parámetro. El área bajo una curva ji cuadrada con ν grados de libertad a la derecha de $\chi^2_{\alpha/2,\nu}$ es $\alpha/2$, lo mismo que a la izquierda de $\chi^2_{1-\alpha/2,\nu}$. De este modo el área capturada entre estos dos valores críticos es $1-\alpha$. Como una consecuencia de esto y del teorema que se acaba de formular,

$$P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) = 1-\alpha \qquad (7.17)$$

Las desigualdades en (7.17) equivalen a

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

Sustituyendo el valor calculado s^2 en los límites se obtiene un intervalo de confianza para σ^2, y sacando las raíces cuadradas se obtiene un intervalo para σ.

Un **intervalo de confianza de 100(1 − α)% para la varianza σ^2 de una población normal** tiene un límite inferior

$$(n-1)s^2/\chi^2_{\alpha/2,n-1}$$

y límite superior

$$(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$$

Un **intervalo de confianza para** σ tiene límites superior e inferior que son las raíces cuadradas de los límites correspondientes en el intervalo para σ^2. Un límite de confianza superior o inferior resulta de la sustitución de $\alpha/2$ con α en el límite correspondiente del intervalo de confianza.

Ejemplo 7.15 Los datos adjuntos sobre voltaje de ruptura de circuitos eléctricamente sobrecargados se tomaron de un diagrama de probabilidad normal que apareció en el artículo "Damage of Flexible Printed Wiring Boards Associated with Lightning-Induced Voltage Surges", (*IEEE Transactions on Components, Hybrids, and Manuf. Tech.*, 1985: 214-220). La derechura del diagrama apoyó de manera firme la suposición de que el voltaje de ruptura está aproximadamente distribuido en forma normal.

1470	1510	1690	1740	1900	2000	2030	2100	2190
2200	2290	2380	2390	2480	2500	2580	2700	

Sea σ^2 la varianza de la distribución del voltaje de ruptura. El valor calculado de la varianza muestral es $s^2 = 137{,}324.3$, la estimación puntual de σ^2. Con grados de libertad $= n - 1 = 16$, un intervalo de confianza de 95% requiere $\chi^2_{.975,16} = 6.908$ y $\chi^2_{.025,16} = 28.845$. El intervalo es

$$\left(\frac{16(137{,}324.3)}{28.845}, \frac{16(137{,}324.3)}{6.908}\right) = (76{,}172.3, 318{,}064.4)$$

Sacando la raíz cuadrada de cada punto extremo se obtiene (276.0, 564.0) como el intervalo de confianza de 95% para σ. Estos intervalos son bastante anchos, lo que refleja la variabilidad sustancial del voltaje de ruptura en combinación con un tamaño de muestra pequeño. ■

Los intervalos de confianza para σ^2 y σ cuando la distribución de la población no es normal pueden ser difíciles de obtener. En esos casos, consulte a un estadístico conocedor.

EJERCICIOS Sección 7.4 (42–46)

42. Determine los valores de las siguientes cantidades:

a. $\chi^2_{.1,15}$ **b.** $\chi^2_{.1,25}$
c. $\chi^2_{.01,25}$ **d.** $\chi^2_{.005,25}$
e. $\chi^2_{.99,25}$ **f.** $\chi^2_{.995,25}$

43. Determine lo siguiente:

a. El 95º percentil de la distribución ji cuadrada con $\nu = 10$
b. El 5º percentil de la distribución ji cuadrada con $\nu = 10$.
c. $P(10.98 \le \chi^2 \le 36.78)$, donde χ^2 es una variable aleatoria ji cuadrada con $\nu = 22$
d. $P(\chi^2 < 14.611 \text{ o } \chi^2 > 37.652)$, donde χ^2 es una variable aleatoria ji cuadrada con $\nu = 25$

44. Se determinó la cantidad de expansión lateral (mils) para una muestra de $n = 9$ soldaduras de arco de gas metálico de energía pulsante utilizadas en tanques de almacenamiento de buques LNG. La desviación estándar muestral resultante fue $s = 2.81$ mils. Suponiendo normalidad, obtenga un intervalo de confianza de 95% para σ^2 y para σ.

45. Se hicieron las siguientes observaciones de tenacidad a la fractura de una placa base de acero maraging con 18% de níquel ["Fracture Testing of Weldments", *ASTM Special Publ. No.* 381, 1965: 328-356 (en kg/pulg^2 $\sqrt{\text{in.}}$, dadas en orden creciente)]:

69.5	71.9	72.6	73.1	73.3	73.5	75.5	75.7
75.8	76.1	76.2	76.2	77.0	77.9	78.1	79.6
79.7	79.9	80.1	82.2	83.7	93.7		

Calcule un intervalo de confianza de 99% para la desviación estándar de la distribución de la tenacidad a la fractura. ¿Es válido este intervalo cualquiera que sea la naturaleza de la distribución? Explique.

46. El artículo "Concrete Pressure on Formwork" (*Mag. of Concrete Res.*, 2009: 407−417) dio las siguientes observaciones sobre la presión máxima del concreto (kN/m^2):

33.2	41.8	37.3	40.2	36.7	39.1	36.2	41.8
36.0	35.2	36.7	38.9	35.8	35.2	40.1	

a. ¿Es factible que esta muestra fuera seleccionada de una distribución de población normal?
b. Calcule un límite de confianza superior con nivel de confianza de 95% para la desviación estándar de la población de presión máxima.

EJERCICIOS SUPLEMENTARIOS (47–62)

47. El ejemplo 1.11 introdujo las observaciones adjuntas sobre fuerza de adhesión.

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

a. Calcule la fuerza de adhesión promedio verdadera de una manera que dé información sobre precisión y confiabilidad. [*Sugerencia*: $\Sigma x_i = 387.8$ y $\Sigma x_i^2 = 4247.08$.]

b. Calcule un intervalo de confianza de 95% para la proporción de todas las adhesiones cuyos valores de fuerza excederían de 10.

48. Un triatlón incluye natación, ciclismo y carrera a pie y es uno de los eventos deportivos amateurs más extenuantes. El artículo "Cardiovascular and Thermal Response of Triathlon Performance" (*Medicine and Science in Sports and Exercise*, 1988: 385–389) reporta sobre un estudio de investigación de nueve triatletas varones. Se registró el ritmo cardiaco máximo (pulsaciones/min) durante la actuación de cada uno de los tres eventos. Para natación, la media y la desviación estándar muestrales fueron 188.0 y 7.2, respectivamente. Suponiendo que la distribución de ritmo cardiaco es (de manera aproximada) normal, construya un intervalo de confianza de 98% para el ritmo cardiaco medio verdadero de triatletas mientras nadan.

49. Para cada uno de los 18 núcleos de depósitos de carbonato humedecidos con aceite, la cantidad de saturación de gas residual después de la inyección de un solvente se midió en la corriente de agua de salida. Las observaciones, en porcentaje de volumen de poros, fueron

23.5	31.5	34.0	46.7	45.6	32.5
41.4	37.2	42.5	46.9	51.5	36.4
44.5	35.7	33.5	39.3	22.0	51.2

(Véase "Relative Permeability Studies of Gas-Water Flow Following Solvent Injection in Carbonate Rocks", *Soc. of Petroleum Engineers J.*, 1976: 23–30.)

a. Construya una gráfica de caja de estos datos y comente sobre cualquier característica interesante.

b. ¿Es factible que la muestra fuera seleccionada de una distribución de población normal?

c. Calcule un intervalo de confianza de 98% para la cantidad promedio verdadera de saturación de gas residual.

50. Un artículo publicado en un periódico reporta que se utilizó una muestra de tamaño 5 como base para calcular un intervalo de confianza de 95% para la frecuencia natural (Hz) promedio verdadera de vigas deslaminadas de cierto tipo. El intervalo resultante fue (229.764, 233.504). Usted decide que un nivel de confianza de 99% es más apropiado que el de 95% utilizado. ¿Cuáles son los límites del intervalo de 99%? [*Sugerencia*: use el centro del intervalo y su ancho para determinar $\bar{x}$ y s.]

51. Una encuesta de 2253 adultos estadounidenses llevada a cabo por el Pew Research Center's Internet & American Life Project en el mes de abril del año 2009 reveló que 1262 de los encuestados había utilizado en algún momento medios inalámbricos para el acceso en línea.

a. Calcule e interprete un intervalo de confianza del 95% para la proporción de todos los adultos estadounidenses que en el momento de la encuesta habían usado medios inalámbricos para el acceso en línea.

b. ¿Qué tamaño de la muestra es necesario si la anchura deseada del intervalo de confianza del 95% debe ser como máximo .04, independientemente de los resultados de la muestra?

c. ¿El límite superior del intervalo en el inciso (a) especifica una fiabilidad del 95% en el límite superior para la proporción calculada? Explique.

52. La alta concentración del elemento tóxico arsénico es demasiado común en el agua subterránea. El artículo "Evaluation of Treatment Systems for the Removal of Arsenic from Groundwater" (*Practice Periodical of Hazardous, Toxic, and Radioactive Waste Mgmt.*, 2005: 152–157) reportó que para una muestra de $n = 5$ especímenes de agua seleccionada para tratamiento por coagulación, la concentración de arsénico media muestral fue de 24.3 μg/L y la desviación estándar muestral fue de 4.1. Los autores del artículo citado utilizaron métodos basados en t para analizar sus datos, así que venturosamente tuvieron razón al creer que la distribución de la concentración de arsénico era normal.

a. Calcule e interprete un intervalo de confianza de 95% para la concentración de arsénico promedio verdadera en todos los especímenes de agua.

b. Calcule un límite de confianza superior de 90% para la desviación estándar de la distribución de la concentración de arsénico.

c. Pronostique la concentración de arsénico de un solo espécimen de agua de modo que dé información sobre precisión y confiabilidad.

53. La infestación con pulgones de árboles frutales puede ser controlada rociando un pesticida o mediante el tratamiento con mariquitas. En un área particular, se seleccionan cuatro diferentes huertas de árboles frutales para experimentación. Las primeras tres arboledas se rocían con los pesticidas 1, 2 y 3, respectivamente, y la cuarta se trata con mariquitas con los siguientes resultados de cosecha:

Tratamiento	$n_i =$ **Número de árboles**	$\bar{x}_i$ **(Bushels/árbol)**	s_i
1	100	10.5	1.5
2	90	10.0	1.3
3	100	10.1	1.8
4	120	10.7	1.6

Sea μ_i = la cosecha promedio verdadera (bushels/árbol) después de recibir el i-ésimo tratamiento. Entonces

$$\theta = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4$$

mide la diferencia de las cosechas promedio verdaderas entre el tratamiento con pesticidas y el tratamiento con mariquitas. Cuando n_1, n_2, n_3 y n_4 son grandes, el estimador $\hat{\theta}$ obtenido al reemplazar cada μ_i con $\overline{X}_i$ es aproximadamente normal. Use esto para deducir un intervalo de confianza muestral grande de $100(1 - \alpha)\%$ para θ y calcule el intervalo de 95% para los datos dados.

54. Es importante que las máscaras utilizadas por bomberos sean capaces de soportar altas temperaturas porque los bomberos comúnmente trabajan en temperaturas de 200–500°F. En una prueba de un tipo de máscara, a 11 de 55 máscaras se les desprendió la mica a 250°. Construya un intervalo de confianza de 90% para la proporción verdadera de máscaras de este tipo cuya mica se desprendería a 250°.

55. Un fabricante de libros de texto universitarios está interesado en investigar la resistencia de las encuadernaciones producidas por una máquina de encuadernar particular. La resistencia puede ser medida registrando la fuerza requerida para arrancar las páginas de la encuadernación. Si esta fuerza se mide en libras, ¿cuántos libros deberán ser probados para calcular la fuerza promedio requerida para romper la encuadernación dentro de .1 lb con 95% de confianza? Suponga que se sabe que σ es de .8.

56. Es bien sabido que la exposición a la fibra de asbesto es un riesgo para la salud. El artículo "The Acute Effects of Chrysotile Asbestos Exposure on Lung Function" (*Environ. Research*, 1978: 360–372) reporta resultados sobre un estudio basado en una muestra de trabajadores de la construcción que habían estado expuestos a asbesto durante un periodo prolongado. Entre los datos dados en el artículo se encontraron los siguientes valores (ordenados) de elasticidad pulmonar ($cm^3/cm\ H_2O$) por cada uno de los 16 sujetos 8 meses después del periodo de exposición (la elasticidad pulmonar mide la elasticidad de los pulmones o cuán efectivamente son capaces de inhalar y exhalar):

167.9	180.8	184.8	189.8	194.8	200.2
201.9	206.9	207.2	208.4	226.3	227.7
228.5	232.4	239.8	258.6		

a. ¿Es factible que la distribución de la población sea normal?
b. Calcule un intervalo de confianza de 95% para la elasticidad pulmonar promedio verdadera después de la exposición.
c. Calcule un intervalo que, con un nivel de confianza de 95%, incluya por lo menos 95% de los valores de elasticidad pulmonar en la distribución de la población.

57. En el ejemplo 6.8, se introdujo el concepto de experimento censurado en el cual n componentes se prueban y el experimento termina en cuanto r de los componentes fallan. Suponga que las vidas útiles de los componentes son independientes, cada uno con distribución exponencial y parámetro λ. Sea Y_1 el tiempo en el cual ocurre la primera falla, Y_2 el tiempo en el cual ocurre la segunda falla, y así sucesivamente, de modo que $T_r = Y_1 + \cdots + Y_r + (n - r)Y_r$, es la vida útil total acumulada. En ese caso se puede demostrar que $2\lambda T_r$ tiene una distribución ji cuadrada con $2r$ grados de libertad. Use esto para desarrollar una fórmula para un intervalo de confianza de $100(1 - \alpha)\%$ para una vida útil promedio verdadera $1/\lambda$. Calcule un intervalo de confianza de 95% con los datos del ejemplo 6.8.

58. Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución de probabilidad continua con mediana $\tilde{\mu}$ (de modo que $P(X_i \le \tilde{\mu}) = P(X_i \ge \tilde{\mu}) = .5$).

a. Demuestre que

$$P(\text{mín}\,(X_i) < \tilde{\mu} < \text{máx}\,(X_i)) = 1 - \left(\frac{1}{2}\right)^{n-1}$$

de modo que (mín(x_i), máx(x_i)) es un intervalo de confianza de $100(1 - \alpha)\%$ para $\tilde{\mu}$ con $\alpha = \left(\frac{1}{2}\right)^{n-1}$. [*Sugerencia*: el complemento del evento $\{\text{mín}\,(X_i) < \tilde{\mu} < \text{máx}\,(X_i)\}$ es $\{\text{máx}\,(X_i) \le \tilde{\mu}\} \cup \{\text{mín}\,(X_i) \ge \tilde{\mu}\}$. Pero máx $(X_i) \le \tilde{\mu}$ si y sólo si $X_i \le \tilde{\mu}$ con todas las i.]

b. Para cada uno de seis infantes normales varones, se determinó la cantidad de alanina aminoácida (mg/100 ml) mientras los infantes llevaban un dieta libre de isoleucina y se obtuvieron los siguientes resultados:

2.84 3.54 2.80 1.44 2.94 2.70

Calcule un intervalo de confianza de 97% para la cantidad mediana verdadera de alanina para infantes que llevaban esa dieta ("The Essential Amino Acid Requirements of Infants", *Amer. J. of Nutrition*, 1964: 322–330).

c. Sea $x_{(2)}$ la segunda más pequeña de las x_i y $x_{(n-1)}$ la segunda más grande de las x_i. ¿Cuál es el coeficiente de confianza del intervalo $(x_{(2)}, x_{(n-1)})$ para $\tilde{\mu}$?

59. Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución uniforme en el intervalo $[0, \theta]$, de modo que

$$f(x) = \begin{cases} \dfrac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{de lo contrario} \end{cases}$$

Entonces si $Y = \text{máx}(X_i)$, se puede demostrar que la variable aleatoria $U = Y/\theta$ tiene una función de densidad

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \le u \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Use $f_U(u)$ para verificar que

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} \le (1 - \alpha/2)^{1/n}\right) = 1 - \alpha$$

y use esto para obtener un intervalo de confianza de $100(1 - \alpha)\%$ para θ.

b. Verifique que $P(\alpha^{1/n} \le Y/\theta \le 1) = 1 - \alpha$ y obtenga un intervalo de confianza de $100(1 - \alpha)\%$ para θ basado en esta proposición de probabilidad.

c. ¿Cuál de los dos intervalos obtenidos previamente es más corto? Si mi tiempo de espera de un camión en la mañana está uniformemente distribuido y los tiempos de espera observados son $x_1 = 4.2$, $x_2 = 3.5$, $x_3 = 1.7$, $x_4 = 1.2$ y $x_5 = 2.4$, obtenga un intervalo de confianza de 95% para θ utilizando el más corto de los dos intervalos.

60. Sea $0 \leq \gamma \leq \alpha$. Entonces un intervalo de confianza de $100(1 - \alpha)\%$ para μ cuando n es grande es

$$\left(\bar{x} - z_\gamma \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha-\gamma} \cdot \frac{s}{\sqrt{n}}\right)$$

La opción $\gamma = \alpha/2$ da el intervalo usual obtenido en la sección 7.2; si $\gamma \neq \alpha/2$, este intervalo no es simétrico con respecto a $\bar{x}$. El ancho de este intervalo es $w = s(z_\gamma + z_{\alpha-\gamma})/\sqrt{n}$. Demuestre que w se reduce al mínimo con la opción $\gamma = \alpha/2$, de modo que el intervalo simétrico sea el más corto. [*Sugerencias*: (a) por definición de z_α, $\Phi(z_\alpha) = 1 - \alpha$, de modo que $z_\alpha = \Phi^{-1}(1 - \alpha)$: (b) la relación entre la derivada de una función $y = f(x)$ y la función inversa $x = f^{-1}(y)$ es $(d/dy)\, f^{-1}(y) = 1/f'(x)$.]

61. Suponga que $x_1, x_2, \ldots, x_n$ son valores observados resultantes de una muestra aleatoria tomada de una distribución simétrica pero posiblemente de cola gruesa. Sean $\tilde{x}$ y f_s la mediana muestral y la dispersión de los cuartos, respectivamente. El capítulo 11 de *Understanding Robust and Exploratory Data Analysis* (véase la bibliografía del capítulo 6) sugiere el siguiente intervalo de confianza de 95% robusto para la media de la población (punto de simetría):

$$\tilde{x} \pm \left(\frac{\text{valor crítico } t \text{ conservador}}{1.075}\right) \cdot \frac{f_s}{\sqrt{n}}$$

El valor de la cantidad entre paréntesis es 2.10 con $n = 10$, 1.94 con $n = 20$ y 1.91 con $n = 30$. Calcule este intervalo de confianza con los datos del ejercicio 45 y compare con el intervalo de confianza t apropiado para una distribución de población normal.

62. **a.** Use los resultados del ejemplo 7.5 para obtener un límite de confianza inferior de 95% para el parámetro λ de una distribución exponencial y calcule el límite basado en los datos dados en el ejemplo.

b. Si la vida útil tiene una distribución exponencial, la probabilidad de que la vida útil exceda de t es $P(X > t) = e^{-\lambda t}$. Use el resultado del inciso (a) para obtener un límite de confianza inferior de 95% para la probabilidad de que el tiempo de ruptura exceda de 100 min.

Bibliografía

DeGroot, Morris y Mark Schervish, *Probability and Statistics* (3a. ed.), Addison-Wesley, Reading, MA, 2002. Una muy buena exposición de los principios generales de inferencia estadística.

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Cengage, Belmont, CA, 2007. La exposición es un poco más completa y sofisticada que la del presente libro e incluye más material sobre bootstrapping.

Hahn, Gerald y William Meeker, *Statistical Intervals*, Wiley, Nueva York, 1991. Todo lo que alguna vez quiso saber sobre intervalos estadísticos (de confianza, predicción, tolerancia y otros).

8 Pruebas de hipótesis basadas en una sola muestra

INTRODUCCIÓN

Un parámetro puede ser estimado a partir de datos muestrales con un solo número (una estimación puntual) o un intervalo completo de valores factibles (un intervalo de confianza). Con frecuencia, sin embargo, el objetivo de una investigación no es estimar un parámetro sino decidir cuál de dos afirmaciones contradictorias sobre el parámetro es la correcta. Los métodos para lograr esto comprenden la parte de la inferencia estadística llamada *prueba de hipótesis*. En este capítulo primero se discuten algunos de los conceptos y terminología básicos en la prueba de hipótesis y luego se desarrollan procedimientos para la toma de decisiones para los problemas de realización de pruebas con base en una muestra tomada de una sola población más frecuentemente encontrados.

8.1 Hipótesis y procedimientos de prueba

Una **hipótesis estadística** o simplemente *hipótesis* es una afirmación o aseveración sobre el valor de un solo parámetro (característica de una población o característica de una distribución de probabilidad), sobre los valores de varios parámetros o sobre la forma de una distribución de probabilidad completa. Un ejemplo de una hipótesis es la pretensión de que $\mu = .75$, donde μ es el diámetro interno promedio verdadero de un cierto tipo de tubo de PVC. Otro ejemplo es la proposición $p < .10$, donde p es la proporción de tarjetas de circuito defectuosas entre todas las tarjetas de circuito producidas por un fabricante. Si μ_1 y μ_2 denotan las resistencias a la ruptura promedio verdaderas de dos tipos diferentes de cuerdas, una hipótesis es la aseveración de que $\mu_1 - \mu_2 = 0$ y otra es que $\mu_1 - \mu_2 > 5$. No obstante, otro ejemplo de una hipótesis es la aseveración de que la distancia de detención en condiciones particulares tiene una distribución normal. Hipótesis de esta última clase se considerarán en el capítulo 14. En éste y en los siguientes capítulos, la atención se concentra en hipótesis en relación con parámetros.

En cualquier problema de prueba de hipótesis, existen dos hipótesis contradictorias en consideración. Una podría ser la pretensión de que $\mu = .75$ y la otra $\mu \neq .75$, o las dos proposiciones contradictorias podrían ser $p \geq .10$ y $p < .10$. El objetivo es decidir, con base en información muestral, cuál de las dos hipótesis es la correcta. Existe una analogía conocida de esto en un juicio criminal. Una pretensión es la aseveración de que el individuo acusado es inocente. En el sistema judicial estadounidense, ésta es la pretensión que inicialmente se cree que es cierta. Sólo de cara a una fuerte evidencia que diga lo contrario el jurado deberá rechazar esta pretensión a favor de la aseveración alternativa de que el acusado es culpable. En este sentido, la pretensión de inocencia es la hipótesis favorecida o protegida y la obligación de la comprobación recae en aquellos que creen en la pretensión alternativa.

Asimismo, al probar hipótesis estadísticas, el problema se formulará de modo que una de las pretensiones sea favorecida al inicio. Esta pretensión inicialmente favorecida no será rechazada a favor de la pretensión alternativa a menos que la evidencia muestral la contradiga y apoye con fuerza la aseveración alternativa.

DEFINICIÓN

La **hipótesis nula** denotada por H_0, es la pretensión que inicialmente se supone cierta (la pretensión de "creencia previa"). La **hipótesis alternativa** denotada por H_a, es la aseveración contradictoria de H_0.

La hipótesis nula será rechazada a favor de la hipótesis alternativa sólo si la evidencia muestral sugiere que H_0 es falsa. Si la muestra no contradice fuertemente a H_0, se continuará creyendo en la factibilidad de la hipótesis nula. Las dos posibles conclusiones derivadas de un análisis de prueba de hipótesis son entonces *rechazar* H_0 o *no rechazar* H_0.

Una **prueba de hipótesis** es un método de utilizar datos muestrales para decidir si la hipótesis nula debe ser rechazada. Por consiguiente se podría probar H_0; $\mu = .75$ contra la H_a alternativa: $\mu \neq .75$. Sólo si los datos muestrales sugieren fuertemente que μ es otra diferente de .75 deberá ser rechazada la hipótesis nula. Sin semejante evidencia, H_0 no deberá ser rechazada, puesto que sigue siendo bastante factible.

En ocasiones un investigador no desea aceptar una aseveración particular a menos y hasta que los datos apoyen fuertemente la aseveración. Como ejemplo, supóngase que una compañía está considerando aplicar un nuevo tipo de recubrimiento en los cojinetes que fabrica. Se sabe que la vida de desgaste promedio verdadera con el recubrimiento actual es de 1000 horas. Si μ denota la vida promedio verdadera del nuevo recubrimiento, la compañía no desea cambiar a menos que la evidencia sugiera fuertemente que μ excede de

1000. Una formulación apropiada del problema implicaría probar H_0: $\mu = 1000$ contra H_a: $\mu > 1000$. La conclusión de que se justifica un cambio está identificada con H_a y se requeriría evidencia conclusiva para justificar el rechazo de H_0 y cambiar al nuevo recubrimiento.

La investigación científica a menudo implica tratar de decidir si una teoría actual debe ser reemplazada por una explicación más factible y satisfactoria del fenómeno investigado. Un método conservador es identificar la teoría actual con H_0 y la explicación alternativa del investigador con H_a. El rechazo de la teoría actual ocurrirá entonces sólo cuando la evidencia sea mucho más compatible con la nueva teoría. En muchas situaciones, H_a se conoce como la "hipótesis del investigador", puesto que es la pretensión que al investigador en realidad le gustaría validar. La palabra *nula* significa "sin ningún valor, efecto o consecuencia", lo que sugiere que H_0 debería ser identificada con la hipótesis de ningún cambio (de la opinión actual), ninguna diferencia, ninguna mejora, y así sucesivamente. Supóngase, por ejemplo, que 10% de todas las tarjetas de circuito producidas por un fabricante durante un periodo reciente estaban defectuosas. Un ingeniero ha sugerido un cambio del proceso de producción en la creencia de que dará por resultado una proporción reducida de tarjetas defectuosas. Sea p la proporción verdadera de tarjetas defectuosas que resultan del proceso cambiado. Entonces la hipótesis de investigación en la cual recae la obligación de la comprobación, es la aseveración de que $p < .10$. Por consiguiente la hipótesis alternativa es H_a: $p < .10$.

En el tratamiento de la prueba de hipótesis, H_0 generalmente será formulada como pretensión de igualdad. Si θ denota el parámetro de interés, la hipótesis nula tendrá la forma H_0: $\theta = \theta_0$ donde θ_0 es un número específico llamado *valor nulo* del parámetro (valor pretendido para θ por la hipótesis nula). Como ejemplo, considérese la situación de la tarjeta de circuito que se acaba de discutir. La hipótesis alternativa sugerida fue H_a: $p < .10$, la pretensión de que la modificación del proceso reduce la proporción de tarjetas defectuosas. Una elección natural de H_0 en esta situación es la pretensión de que $p \geq .10$ de acuerdo con la cual el nuevo proceso no es mejor *o* peor que el actualmente utilizado. En su lugar se considerará H_0: $p = .10$ contra H_a: $p < .10$. El razonamiento para utilizar esta hipótesis nula simplificada es que cualquier procedimiento de decisión razonable para decidir entre H_0: $p = .10$ y H_a: $p < .10$ también será razonable para decidir entre la pretensión de que $p \geq .10$ y H_a. Se prefiere utilizar una H_0 simplificada porque tiene ciertos beneficios técnicos, los que en breve serán aparentes.

La alternativa a la hipótesis nula H_0: $\theta = \theta_0$ se verá como una de las siguientes tres aseveraciones:

1. H_a: $\theta > \theta_0$ (en cuyo caso la hipótesis nula implícita es $\theta \leq \theta_0$),

2. H_a: $\theta < \theta_0$ (en cuyo caso la hipótesis nula implícita es $\theta \geq \theta_0$), o

3. H_a: $\theta \neq \theta_0$

Por ejemplo, sea σ la desviación estándar de la distribución de diámetros internos (pulgadas) de cierto tipo de manguito de metal. Si se decidió utilizar el manguito a menos que la evidencia muestral demuestre conclusivamente que $\sigma > .001$, la hipótesis apropiada sería H_0: $\sigma = .001$ contra H_a: $\sigma > .001$. El número θ_0 que aparece tanto en H_0 como en H_a (separando la alternativa de la nula) se llama **valor nulo.**

Procedimientos de prueba

Un procedimiento de prueba es una regla, basada en datos muestrales, para decidir si se rechaza H_0. Una prueba de H_0: $p = .10$ contra H_a: $p < .10$ en el problema de tarjetas de circuito podría estar basada en examinar una muestra aleatoria de $n = 200$ tarjetas. Sea X el número de tarjetas defectuosas en la muestra, una variable aleatoria binomial; x representa el valor observado de X. Si H_0 es verdadera, $E(X) = np = 200(.10) = 20$, en tanto que se pueden esperar menos de 20 tarjetas defectuosas si H_a es verdadera. Un valor de x

un poco por debajo de 20 no contradice fuertemente a H_0, así que es razonable rechazar H_0 sólo si x es de manera sustancial menor que 20. Un procedimiento de prueba como ése es rechazar H_0 si $x \leq 15$ y no rechazarla de lo contrario. Este procedimiento consta de dos constituyentes: (1) un *estadístico de prueba* o función de los datos muestrales utilizados para tomar la decisión y (2) una *región de rechazo* compuesta de aquellos valores x con los cuales H_0 será rechazada a favor de H_a. De acuerdo con la regla que se acaba de sugerir, la región de rechazo se compone de $x = 0, 1, 2, \ldots,$ y 15. H_0 no será rechazada si $x = 16, 17, \ldots, 199$ o 200.

Un procedimiento de prueba se especifica como sigue:

1. Un **estadístico de prueba,** una función de los datos muestrales en los cuales ha de basarse la decisión (rechazar H_0 o no rechazar H_0)
2. Una **región de rechazo,** el conjunto de todos los valores estadísticos de prueba por los cuales H_0 será rechazada

La hipótesis nula será rechazada entonces si y sólo si el valor estadístico de prueba observado o calculado queda en la región de rechazo.

Como otro ejemplo, supóngase que una compañía tabacalera afirma que el contenido de nicotina promedio μ de los cigarrillos marca B es (cuando mucho) de 1.5 mg. Sería imprudente rechazar la afirmación del fabricante sin una fuerte evidencia contradictoria, así que una formulación apropiada del problema es probar H_0: $\mu = 1.5$ contra H_a: $\mu > 1.5$. Considérese una regla de decisión basada en analizar una muestra aleatoria de 32 cigarrillos. Sea $\overline{X}$ el contenido de nicotina promedio muestral. Si H_0 es verdadera $E(\overline{X}) = \mu = 1.5$ en tanto que si H_0 es falsa, se espera que $\overline{X}$ exceda de 1.5. Una fuerte evidencia contra H_0 es proporcionada por un valor de $\overline{x}$ que exceda considerablemente de 1.5. Por consiguiente, se podría utilizar $\overline{X}$ como un estadístico de prueba junto con la región de rechazo $\overline{x} \geq 1.6$.

Tanto en el ejemplo de tarjetas de circuito como en el de contenido de nicotina, la selección del estadístico de prueba y la forma de la región de rechazo tienen sentido intuitivamente. Sin embargo, la selección del valor de corte utilizado para especificar la región de rechazo es un tanto arbitraria. En lugar de rechazar H_0: $p = .10$ a favor de H_a: $p < .10$ cuando $x \leq 15$, se podría utilizar la región de rechazo $x \leq 14$. En esta región, H_0 no sería rechazada si se observaran 15 tarjetas defectuosas, mientras que esta ocurrencia conduciría al rechazo de H_0 si se emplea la región inicialmente sugerida. Asimismo, se podría utilizar la región de rechazo $\overline{x} \geq 1.55$ en el problema de contenido de nicotina en lugar de la región $\overline{x} \geq 1.60$.

Errores en la prueba de hipótesis

La base para elegir una región de rechazo particular radica en la consideración de los errores que se podrían presentar al sacar una conclusión. Considérese la región de rechazo $x \leq 15$ en el problema de tarjetas de circuito. Aun cuando H_0: $p = .10$ sea verdadera, podría suceder que una muestra inusual dé por resultado $x = 13$, de modo que H_0 sea erróneamente rechazada. Por otra parte, aun cuando H_a: $p < .10$ sea verdadera, una muestra inusual podría dar $x = 20$, en cuyo caso H_0 no sería rechazada, de nueva cuenta una conclusión incorrecta. Por lo tanto, es posible que H_0 pueda ser rechazada cuando sea verdadera o que H_0 no pueda ser rechazada cuando sea falsa. Estos posibles errores no son consecuencias de una región de rechazo imprudentemente seleccionada. Cualquiera de estos dos errores podría presentarse cuando se emplea la región $x \leq 14$, o cuando se utiliza cualquier otra región sensible.

DEFINICIÓN

Un **error de tipo I** consiste en rechazar la hipótesis nula H_0 cuando es verdadera. Un **error de tipo II** implica no rechazar H_0 cuando H_0 es falsa.

En el problema de contenido de nicotina, un error de tipo I consiste en rechazar la afirmación del fabricante de que $\mu = 1.5$ cuando en realidad es cierta. Si se emplea la región de rechazo $\bar{x} \geq 1.6$, podría suceder que $\bar{x} = 1.63$ aun cuando $\mu = 1.5$, con el resultado de un error de tipo I. Alternativamente, puede ser que H_0 sea falsa y no obstante sea observada $\bar{x} = 1.52$, lo que conduciría a que H_0 no sea rechazada (un error de tipo II).

En el mejor de todos los mundos posibles, se podrían desarrollar procedimientos de prueba en los cuales ningún tipo de error es posible. Sin embargo, este ideal puede ser alcanzado sólo si la decisión se basa en el examen de toda la población. La dificultad con la utilización de un procedimiento basado en datos muestrales es que debido a la variabilidad del muestreo, el resultado podría ser una muestra no representativa, es decir, un valor de $\bar{X}$ que está lejos de μ o un valor de $\hat{p}$ que difiere considerablemente de p.

En lugar de demandar procedimientos sin errores, habrá que buscar procedimientos con los cuales sea improbable que ocurra cualquier tipo de error. Es decir, un buen procedimiento es uno con el cual la probabilidad de cometer cualquier tipo de error es pequeña. La selección de un valor de corte en una región de rechazo particular fija las probabilidades de errores de tipo I y tipo II. Estas probabilidades de error son tradicionalmente denotadas por α y β, de manera respectiva. Como H_0 especifica un valor único del parámetro, existe un solo valor de α. Sin embargo, existe un valor diferente de β para cada valor del parámetro compatible con H_a.

Ejemplo 8.1 Se sabe que cierto tipo de automóvil no sufre daños visibles el 25% del tiempo en pruebas de choque a 10 mph. Se ha propuesto un diseño de parachoques modificado en un esfuerzo por incrementar este porcentaje. Sea p la proporción de todos los choques a 10 mph con este nuevo parachoques en los que no se producen daños visibles. Las hipótesis a ser tratadas son H_0: $p = .25$ (ninguna mejora) contra H_a: $p > .25$. La prueba se basará en un experimento que implica $n = 20$ choques independientes con prototipos del nuevo diseño. Intuitivamente, H_0 deberá ser rechazada si un número sustancial de los choques no muestra daños. Considérese el siguiente procedimiento de prueba:

Estadístico de prueba: X = número de choques sin daños visibles

Región de rechazo: $R_8 = \{8, 9, 10, \ldots, 19, 20\}$; es decir, rechazar H_0 si $x \geq 8$, donde x es el valor observado del estadístico de prueba.

Esta región de rechazo se llama *de cola superior* porque se compone de sólo grandes valores del estadístico de prueba.

Cuando H_0 es verdadera, la distribución de probabilidad de X es binomial con $n = 20$ y $p = .25$. Entonces

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera})\\ &= P(X \geq 8 \text{ cuando } X \sim \text{Bin}(20, .25)) = 1 - B(7; 20, .25)\\ &= 1 - .898 = .102\end{aligned}$$

Es decir, cuando H_0 en realidad es verdadera, aproximadamente 10% de todos los experimentos compuestos de 20 choques darían por resultado que H_0 fuera rechazada incorrectamente (un error de tipo I).

En contraste con α, no hay una sola β. En su lugar, hay una β diferente por cada p distinta que exceda de .25. Por consiguiente, hay un valor de β con p =.3 (en cuyo caso $X \sim \text{Bin}(20, .3)$), otro valor de β con $p = .5$ y así sucesivamente. Por ejemplo,

$$\begin{aligned}\beta(.3) &= P(\text{error de tipo II cuando } p = .3)\\ &= P(H_0 \text{ no es rechazada cuando es falsa porque } p = .3)\\ &= P(X \leq 7 \text{ cuando } X \sim \text{Bin}(20, .3)) = B(7; 20, .3) = .772\end{aligned}$$

Cuando p es en realidad .3 y no .25 (un "pequeño" alejamiento de H_0), ¡aproximadamente 77% de todos los experimentos de este tipo darían por resultado que H_0 no sea rechazada de manera incorrecta!

La tabla adjunta muestra β para valores seleccionados de p (cada uno calculado para la región de rechazo R_8). Claramente, β disminuye conforme el valor de p se aleja hacia la derecha del valor nulo .25. De manera intuitiva, mientras más grande es el alejamiento de H_0, menos probable es que dicho alejamiento no sea detectado.

p	.3	.4	.5	.6	.7	.8
$\beta(p)$	.772	.416	.132	.021	.001	.000

El procedimiento de prueba propuesto sigue siendo razonable para poner a prueba la hipótesis nula más realista de que $p \le .25$. En este caso, ya no existe una sola α, sino que hay una α por cada p que sea cuando mucho de .25: $\alpha(.25)$, $\alpha(.23)$, $\alpha(.20)$, $\alpha(.15)$ y así sucesivamente. Es fácil verificar, no obstante, que $\alpha(p) < \alpha(.25) = .102$ si $p < .25$. Es decir, el valor más grande de α ocurre con el valor límite .25 entre H_0 y H_a. Por consiguiente si α es pequeña para la hipótesis nula simplificada, también será igual o más pequeña para la H_0 más realista. ■

Ejemplo 8.2 Se sabe que el tiempo de secado de un tipo de pintura en condiciones de prueba especificadas está normalmente distribuido con valor medio de 75 min y desviación estándar de 9 min. Algunos químicos propusieron un nuevo aditivo para reducir el promedio de tiempo de secado. Se cree que los tiempos de secado con este aditivo permanecerán distribuidos en forma normal con $\sigma = 9$. Debido al gasto asociado con el aditivo, la evidencia deberá sugerir fuertemente una mejora en el tiempo de secado promedio antes de que se adopte semejante conclusión. Sea μ el tiempo de secado promedio verdadero cuando se utiliza el aditivo. Las hipótesis apropiadas son H_0; $\mu = 75$ contra H_a: $\mu < 75$. Sólo si H_0 puede ser rechazada el aditivo será declarado exitoso y utilizado.

Los datos experimentales tienen que estar compuestos de tiempos de secado de $n = 25$ especímenes de prueba. Sean $X_1, \ldots, X_{25}$ los 25 tiempos de secado, una muestra aleatoria de tamaño 25 de una distribución normal con valor medio μ y desviación estándar $\sigma = 9$. El tiempo de secado medio muestral $\overline{X}$ tiene entonces una distribución normal con valor esperado $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 9/\sqrt{25} = 1.80$. Cuando H_0 es verdadera, $\mu_{\overline{X}} = 75$, así que un valor $\bar{x}$ sustancialmente menor que 75 contradiría fuertemente a H_0. Una región razonable de rechazo tiene la forma $\bar{x} \le c$, donde el valor de corte c es adecuadamente seleccionado. Considere la opción $c = 70.8$, de modo que el procedimiento de prueba se componga del estadístico de prueba $\overline{X}$ y una región de rechazo $\bar{x} \le 70.8$. Debido a que la región de rechazo se compone de sólo valores pequeños del estadístico de prueba, se dice que ésta es *de cola pequeña*. El cálculo de α y β ahora implica una estandarización de rutina de $\overline{X}$ seguida por una referencia a las probabilidades normales estándar de la tabla A.3 del apéndice:

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera})\\ &= P(\overline{X} \le 70.8 \text{ cuando } \overline{X} \sim \text{normal con } \mu_{\overline{X}} = 75, \sigma_{\overline{X}} = 1.8)\\ &= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = .01\end{aligned}$$

$$\begin{aligned}\beta(72) &= P(\text{error de tipo II cuando } \mu = 72)\\ &= P(H_0 \text{ no es rechazada cuando es falsa porque } \mu = 72)\\ &= P(\overline{X} > 70.8 \text{ cuando } \overline{X} \sim \text{normal con } \mu_{\overline{X}} = 72 \text{ y } \sigma_{\overline{X}} = 1.8)\\ &= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) = 1 - \Phi(-.67) = 1 - .2514 = .7486\end{aligned}$$

$$\beta(70) = 1 - \Phi\left(\frac{70.8 - 70}{1.8}\right) = .3300 \qquad \beta(67) = .0174$$

Para el procedimiento de prueba especificado, sólo 1% de todos los experimentos realizados como se describió darán por resultado que H_0 sea rechazada cuando en realidad es verdadera. No obstante, la probabilidad de un error de tipo II es muy grande cuando $\mu = 72$ (sólo un pequeño alejamiento de H_0), un poco menor cuando $\mu = 70$ y bastante pequeño cuando $\mu = 67$ (un alejamiento muy sustancial de H_0). Estas probabilidades de error se ilustran en la figura 8.1. Obsérvese que α se calcula con la distribución de probabilidad del estadístico de prueba cuando H_0 es verdadera, en tanto que la determinación de β requiere conocer la distribución del estadístico de prueba cuando H_0 es falsa.

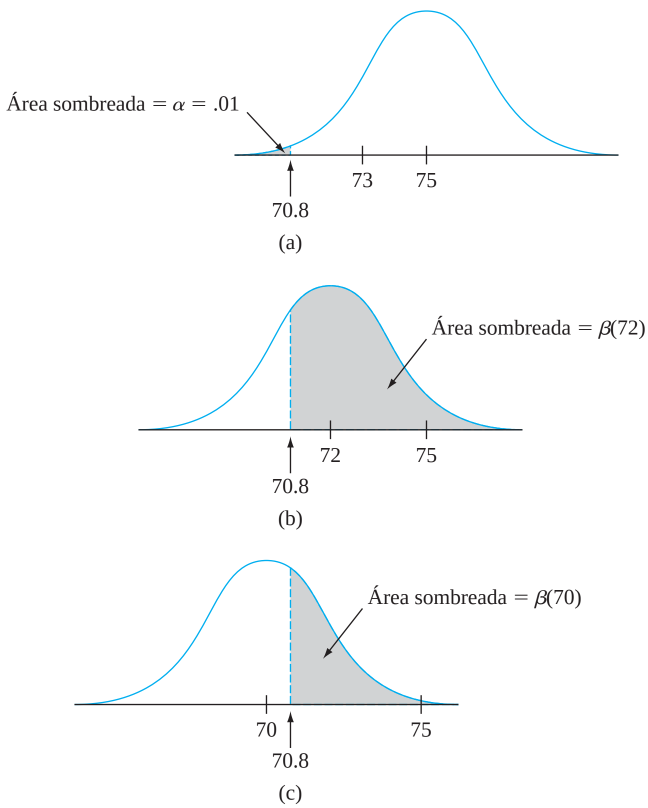


Figura 8.1 α y β ilustradas para el ejemplo 8.2: (a) la distribución de $\overline{X}$ cuando $\mu = 75$ (H_0 verdadera); (b) la distribución de $\overline{X}$ cuando $\mu = 72$ (H_0 falsa); (c) la distribución de $\overline{X}$ cuando $\mu = 70$ (H_0 falsa)

Como en el ejemplo 8.1, si se considera la hipótesis nula más realista $\mu \geq 75$, existe una α para cada valor de parámetro con el cual H_0 es verdadera: $\alpha(75)$, $\alpha(75.8)$, $\alpha(76.5)$, y así de manera sucesiva. Es fácil verificar que $\alpha(75)$ es la más grande de todas estas probabilidades de error de tipo I. Enfocar en el valor límite equivale a trabajar explícitamente con el "peor caso". ■

La especificación de un valor de corte para la región de rechazo en los ejemplos que se acaban de considerar fue algo arbitraria. El uso de $R_8\{8, 9, \ldots, 20\}$ en el ejemplo 8.1 dio por resultado $\alpha = .102$, $\beta(.3) = .772$ y $\beta(.5) = .132$. Muchos pensarán que estas probabilidades de error son intolerablemente grandes. Quizás puedan reducirse si se cambia el valor de corte.

Ejemplo 8.3 (Continuación del ejemplo 8.1)

Utilice el mismo experimento y el estadístico de prueba X como previamente se describió en el problema de la defensa de automóvil pero ahora considere la región de rechazo $R_9 = \{9, 10, \ldots, 20\}$. Como X sigue teniendo una distribución binomial con parámetros $n = 20$ y p,

$$\begin{aligned}\alpha &= P(H_0 \text{ es rechazada cuando } p = .25)\\ &= P(X \geq 9 \text{ cuando } X \sim \text{Bin}(20, .25)) = 1 - B(8; 20, .25) = .041\end{aligned}$$

La probabilidad de error de tipo I se redujo con el uso de la nueva región de rechazo. Sin embargo, se pagó un precio por esta reducción:

$$\begin{aligned}\beta(.3) &= P(H_0 \text{ no es rechazada cuando } p = .3)\\ &= P(X \leq 8 \text{ cuando } X \sim \text{Bin}(20, .3)) = B(8; 20, .3) = .887\\ \beta(.5) &= B(8; 20, .5) = .252\end{aligned}$$

Ambas β son más grandes que las probabilidades de error correspondientes .772 y .132 para la región R_8. En retrospectiva, esto no es sorprendente; α se calcula sumando las probabilidades de los valores estadísticos de prueba *en la región de rechazo*, en tanto que β es la probabilidad de que X quede *en el complemento* de la región de rechazo. Al hacerse más pequeña la región de rechazo debe reducirse α al mismo tiempo que se incrementa β con cualquier $p > .25$. ■

Ejemplo 8.4 (Continuación del ejemplo 8.2)

El uso del valor de corte $c = 70.8$ en el ejemplo de secado de la pintura dio por resultado un valor de α muy pequeño (.01) pero β grandes. Considere el mismo experimento y el estadístico de prueba $\overline{X}$ con la nueva región de rechazo $\overline{x} \leq 72$. Como $\overline{X}$ sigue siendo normalmente distribuida con valor medio $\mu_{\overline{X}} = \mu$ y $\sigma_{\overline{X}} = 1.8$,

$$\begin{aligned}\alpha &= P(H_0 \text{ es rechazada cuando es verdadera})\\ &= P(\overline{X} \leq 72 \text{ cuando } \overline{X} \sim N(75, 1.8^2))\\ &= \Phi\left(\frac{72 - 75}{1.8}\right) = \Phi(-1.67) = .0475 \approx .05\end{aligned}$$

$$\begin{aligned}\beta(72) &= P(H_0 \text{ no es rechazada cuando } \mu = 72)\\ &= P(\overline{X} > 72 \text{ cuando } \overline{X} \text{ es una variable aleatoria normal con media de 72 y desviación estándar de 1.8})\\ &= 1 - \Phi\left(\frac{72 - 72}{1.8}\right) = 1 - \Phi(0) = .5\\ \beta(70) &= 1 - \Phi\left(\frac{72 - 70}{1.8}\right) = .1335 \qquad \beta(67) = .0027\end{aligned}$$

El cambio del valor de corte agrandó la región de rechazo (incluye más valores de $\overline{x}$) y el resultado es una reducción de β por cada μ fija menor que 75. Sin embargo, α en esta nueva región se ha incrementado desde el valor previo .01 hasta aproximadamente .05. Si una probabilidad de error de tipo I así de grande puede ser tolerada, se prefiere la segunda región ($c = 72$) a la primera ($c = 70.8$) debido a las β más pequeñas. ■

Los resultados de estos ejemplos pueden ser generalizados de la siguiente manera.

PROPOSICIÓN

> Supóngase que un experimento y un tamaño de muestra están fijos y que se selecciona un estadístico de prueba. Entonces si se reduce el tamaño de la región de rechazo para obtener un valor más pequeño de α se obtiene un valor más grande de β con cualquier valor de parámetro particular compatible con H_a.

Esta proposición expresa que una vez que el estadístico de prueba y n están fijos, no existe una región de rechazo que haga que al mismo tiempo α y β sean pequeños. Se debe seleccionar una región para establecer un compromiso entre α y β.

Debido a las indicaciones sugeridas para especificar H_0 y H_a, casi siempre un error de tipo I es más serio que uno de tipo II (esto en general se puede lograr mediante la selección apropiada de las hipótesis). El método seguido por la mayoría de los practicantes de la estadística es especificar el valor más grande de α que pueda ser tolerado y encontrar una región de rechazo que tenga ese valor de α en lugar de cualquier otro más pequeño. Esto hace a β tan pequeño como sea posible sujeto al límite en α. El valor resultante de α

a menudo se conoce como **nivel de significancia** de la prueba. Niveles tradicionales de significancia son .10, .05 y .01, aunque el nivel en cualquier problema particular dependerá de la seriedad de un error de tipo I; mientras más serio es este error, más pequeño deberá ser el nivel de significancia. El procedimiento de prueba correspondiente se llama **prueba de nivel α** (p. ej., prueba de nivel .05 o prueba de nivel .01). Una prueba con nivel de significancia α es una donde la probabilidad de error de tipo I se controla al nivel especificado.

Ejemplo 8.5 Nuevamente sea μ que denota el contenido de nicotina promedio verdadero de los cigarrillos marca B. El objetivo es probar H_0: $\mu = 1.5$ contra H_a: $\mu > 1.5$ con base en una muestra aleatoria $X_1, X_2, \ldots, X_{32}$ de contenido de nicotina. Suponga que se sabe que la distribución del contenido de nicotina es normal con $\sigma = .20$. Entonces $\overline{X}$ está normalmente distribuida con valor medio $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = .20/\sqrt{32} = .0354$.

En lugar de utilizar $\overline{X}$ como estadístico de prueba, se estandariza $\overline{X}$ suponiendo que H_0 es verdadera.

$$\text{Estadístico de prueba: } Z = \frac{\overline{X} - 1.5}{\sigma/\sqrt{n}} = \frac{\overline{X} - 1.5}{.0354}$$

Z expresa la distancia entre $\overline{X}$ y su valor esperado cuando H_0 es verdadera como algún número de desviaciones estándar. Por ejemplo, $z = 3$ resulta de una $\overline{x}$ que es 3 desviaciones estándar más grande de lo que se habría esperado si H_0 fuera verdadera.

Rechazar H_0 cuando $\overline{x}$ excede "considerablemente" de 1.5 equivale a rechazar H_0 cuando z excede "considerablemente" de cero. Es decir, la forma de la región de rechazo es $z \geq c$. Determínese ahora c de modo que $\alpha = .05$. Cuando H_0 es verdadera, Z tiene una distribución estándar normal. Por consiguiente

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera})\\ &= P(Z \geq c \text{ cuando } Z \sim N(0, 1))\end{aligned}$$

El valor de c debe capturar el área de la cola superior .05 bajo la curva z. De la sección 4.3 o directamente de la tabla A.3 del apéndice, $c = z_{.05} = 1.645$.

Obsérvese que $z \geq 1.645$ equivale a $\overline{x} - 1.5 \geq (.0354)(1.645)$, es decir, $\overline{x} \geq 1.56$. Entonces β implica la probabilidad de que $\overline{X} < 1.56$ y puede ser calculada para cualquier μ mayor que 1.5. ■

EJERCICIOS Sección 8.1 (1–14)

1. Por cada una de las siguientes aseveraciones, exprese si es una hipótesis estadística legítima y por qué:
- **a.** H: $\sigma > 100$ **b.** H: $\tilde{x} = 45$
- **c.** H: $s \leq .20$ **d.** H: $\sigma_1/\sigma_2 < 1$
- **e.** H: $\overline{X} - \overline{Y} = 5$
- **f.** H: $\lambda \leq .01$ donde λ es el parámetro de una distribución exponencial utilizada para modelar la vida útil de un componente

2. Para los siguientes pares de aseveraciones, indique cuáles no satisfacen las reglas para establecer hipótesis y por qué (los subíndices 1 y 2 diferencian las cantidades para dos poblaciones o muestras diferentes).
- **a.** H_0: $\mu = 100$, H_a: $\mu > 100$
- **b.** H_0: $\sigma = 20$, H_a: $\sigma \leq 20$
- **c.** H_0: $p \neq .25$, H_a: $p = .25$
- **d.** H_0: $\mu_1 - \mu_2 = 25$, H_a: $\mu_1 - \mu_2 > 100$
- **e.** H_0: $S_1^2 = S_2^2$, H_a: $S_1^2 \neq S_2^2$
- **f.** H_0: $\mu = 120$, H_a: $\mu = 150$
- **g.** H_0: $\sigma_1/\sigma_2 = 1$, H_a: $\sigma_1/\sigma_2 \neq 1$
- **h.** H_0: $p_1 - p_2 = -.1$, H_a: $p_1 - p_2 < -.1$

3. Para determinar si las soldaduras de las tuberías en una planta de energía nuclear satisfacen las especificaciones, se selecciona una muestra aleatoria de soldaduras y se realizan pruebas en cada una de ellas. La resistencia de la soldadura se mide como la fuerza requerida para romperla. Suponga que las especificaciones indican que la resistencia media de las soldaduras deberá exceder de 100 lb/pulg^2; el equipo de inspección decide probar H_0: $\mu = 100$ contra H_a: $\mu > 100$. Explique por qué podría ser preferible utilizar esta H_a en lugar de $\mu < 100$.

4. Sea μ el nivel de radiactividad promedio verdadero (picocuries por litro). Se considera que el valor 5 pCi/L es la línea divisoria entre agua segura e insegura. ¿Recomendaría probar H_0: $\mu = 5$ contra H_a: $\mu > 5$ o H_0: $\mu = 5$ contra H_a: $\mu < 5$? Explique su razonamiento [*Sugerencia*: piense en las consecuencias de un error de tipo I o de un error de tipo II con cada posibilidad.]

5. Antes de aprobar la compra de un gran pedido de fundas de polietileno para un tipo particular de cable de energía submarino relleno de aceite a alta presión, una compañía desea contar con evidencia conclusiva de que la desviación estándar verdadera del espesor de la funda es de menos de .05 mm. ¿Qué hipótesis deberán ser probadas y por qué? En este contexto, ¿cuáles son los errores de tipos I y II?

6. Muchas casas viejas cuentan con sistemas eléctricos que utilizan fusibles en lugar de interruptores de circuito. Un fabricante de fusibles de 40 amp desea asegurarse de que el amperaje medio al cual se queman sus fusibles es en realidad de 40. Si el amperaje medio es menor que 40, los clientes se quejarán porque los fusibles tienen que ser reemplazados con demasiada frecuencia. Si el amperaje medio es de más de 40, el fabricante podría ser responsable de los daños que sufra un sistema eléctrico a causa del funcionamiento defectuoso de los fusibles. Para verificar el amperaje de los fusibles, se selecciona e inspecciona una muestra de fusibles. Si tuviera que realizarse un prueba de hipótesis con los datos resultantes, ¿qué hipótesis nula y alternativa serían de interés para el fabricante? Describa los errores de tipos I y II en el contexto de este problema.

7. Se toman muestras de agua utilizada para enfriamiento al momento de ser descargada por una planta de energía en un río. Se ha determinado que en tanto la temperatura media del agua descargada sea cuando mucho de 150°F, no habrá efectos negativos en el ecosistema del río. Para investigar si la planta cumple con los reglamentos que prohíben una temperatura media por encima de 150° del agua de descarga, se tomarán al azar 50 muestras de agua y se registrará la temperatura de cada una. Los datos resultantes se utilizarán para probar la hipótesis H_0: $\mu = 150°$ contra H_a: $\mu > 150°$. En el contexto de esta situación, describa los errores de tipo I y tipo II. ¿Qué tipo de error consideraría más serio? Explique.

8. Un tipo regular de laminado está siendo utilizado por un fabricante de tarjetas de circuito. Un laminado especial ha sido desarrollado para reducir la combadura. El laminado regular será utilizado en una muestra de especímenes y el laminado especial en otra muestra y se determinará entonces la cantidad de combadura en cada espécimen. El fabricante cambiará entonces al laminado especial sólo si puede demostrar que la cantidad de combadura promedio verdadera de dicho laminado es menor que la del laminado regular. Formule las hipótesis pertinentes y describa los errores de tipo I y de tipo II en el contexto de esta situación.

9. Dos compañías diferentes han solicitado proporcionar el servicio de televisión por cable en una región. Sea p la proporción de todos los suscriptores potenciales que favorecen a la primera compañía sobre la segunda. Considere probar H_0: $p = .5$ contra H_a: $p \neq .5$ basado en una muestra aleatoria de 25 individuos. Sea X el número en la muestra que favorecen a la primera compañía y x el valor observado de X.

a. ¿Cuál de las siguientes regiones de rechazo es más apropiada y por qué?

$$R_1 = \{x: x \leq 7 \text{ o } x \geq 18\}, R_2 = \{x: x \leq 8\},$$
$$R_3 = \{x: x \geq 17\}$$

b. En el contexto de este problema, describa cuáles son los errores de tipo I y de tipo II.

c. ¿Cuál es la distribución de probabilidad del estadístico de prueba X cuando H_0 es verdadera? Úsela para calcular la probabilidad de un error de tipo I.

d. Calcule la probabilidad de un error de tipo II en la región seleccionada cuando $p = .3$, otra vez cuando $p = .4$ y también con $p = .6$ y $p = .7$.

e. Utilizando la región seleccionada, ¿qué concluiría si 6 de los 25 individuos encuestados favorecen a la compañía 1?

10. Una mezcla de cenizas combustibles pulverizadas y cemento Portland utilizada para rellenar con lechada deberá tener una resistencia a la compresión de más de 1300 KN/m^2. La mezcla no será utilizada a menos que la evidencia experimental indique concluyentemente que la especificación de resistencia ha sido satisfecha. Suponga que la resistencia a la compresión de especímenes de esta muestra está normalmente distribuida con $\sigma = 60$. Sea μ la resistencia a la compresión promedio verdadera.

a. ¿Cuáles son las hipótesis nula y alternativa apropiadas?

b. Sea $\overline{X}$ la resistencia a la compresión promedio muestral de $n = 10$ especímenes seleccionados al azar. Considere el procedimiento de prueba con estadístico de prueba $\overline{X}$ y región de rechazo $\overline{x} \geq 1331.26$. ¿Cuál es la distribución de probabilidad del estadístico de prueba cuando H_0 es verdadera? ¿Cuál es la probabilidad de un error de tipo I para el procedimiento de prueba?

c. ¿Cuál es la distribución de probabilidad del estadístico de prueba cuando $\mu = 1350$? Utilizando el procedimiento de prueba del inciso (b), ¿cuál es la probabilidad de que la mezcla sea juzgada insatisfactoria cuando en realidad $\mu = 1350$ (un error de tipo II)?

d. ¿Cómo cambiaría el procedimiento de prueba del inciso (b) para obtener una prueba con nivel de significancia de .05? ¿Qué impacto tendría este cambio en la probabilidad de error del inciso (c)?

e. Considere el estadístico de prueba estandarizado $Z = (\overline{X} - 1300)/(\sigma/\sqrt{n}) = (\overline{X} - 1300)/13.42$. ¿Cuáles son los valores de Z correspondientes a la región de rechazo del inciso (b)?

11. La calibración de una báscula tiene que ser verificada pesando 25 veces un espécimen de prueba de 10 kg. Suponga que los resultados de diferentes pesadas son independientes entre sí y que el peso en cada ensayo está normalmente distribuido con $\sigma = .200$ kg. Sea μ la lectura de peso promedio verdadero en la báscula.

a. ¿Qué hipótesis deberá ponerse a prueba?

b. Suponga que la báscula tiene que ser recalibrada si $\overline{x} \geq 10.1032$ o $\overline{x} \leq 9.8968$. ¿Cuál es la probabilidad de que se realice la recalibración cuando en realidad no es necesaria?

c. ¿Cuál es la probabilidad de que la recalibración sea considerada innecesaria cuando en realidad $\mu = 10.1$? ¿Cuando $\mu = 9.8$?

d. Sea $z = (\overline{x} - 10)/(\sigma/\sqrt{n})$. ¿Con qué valor de c la región de rechazo del inciso (b) equivale a la región de "dos colas" $z \geq c$ o $z \leq -c$?

e. Si el tamaño de muestra fuera de sólo 10 y no de 25, ¿cómo modificaría el procedimiento del inciso (d) de modo que $\alpha = .05$?

f. Utilizando la prueba del inciso (e), ¿qué concluiría basado en los siguientes datos muestrales?

9.981	10.006	9.857	10.107	9.888
9.728	10.439	10.214	10.190	9.793

g. Exprese de nuevo el procedimiento de prueba del inciso (b) en función del estadístico de prueba estandarizado $Z = (\overline{X} - 10)/(\sigma/\sqrt{n})$.

12. Se ha propuesto un nuevo diseño del sistema de frenos de un tipo de carro. Para el sistema actual, se sabe que la distancia de frenado promedio verdadera a 40 mph en condiciones específicas es de 120 pies. Se propone que el nuevo diseño sea instalado sólo si los datos muestrales indican fuertemente una reducción de la distancia de frenado promedio verdadera del nuevo diseño.

a. Defina el parámetro de interés y formule las hipótesis pertinentes.

b. Suponga que la distancia de frenado del nuevo sistema está normalmente distribuida con $\sigma = 10$. Sea $\overline{X}$ la distancia de frenado promedio de una muestra aleatoria de 36 observaciones. ¿Cuál de las siguientes tres regiones de rechazo es apropiada: $R_1 = \{\bar{x}: \bar{x} \geq 124.80\}$, $R_2 = \{\bar{x}: \bar{x} \leq 115.20\}$, $R_3 = \{\bar{x}: \bar{x} \geq 125.13$ o $\bar{x} \leq 114.87\}$?

c. ¿Cuál es el nivel de significancia de la región apropiada del inciso (b)? ¿Cómo cambiaría la región para obtener una prueba con $\alpha = .001$?

d. ¿Cuál es la probabilidad de que el nuevo diseño no sea instalado cuando la distancia de frenado promedio verdadera sea en realidad de 115 pies y la región apropiada del inciso (b) sea utilizada?

e. Sea $Z = (\overline{X} - 120)/(\sigma/\sqrt{n})$. ¿Cuál es el nivel de significancia de la región de rechazo $\{z: z \leq -2.33\}$? ¿Y para la región $\{z: z \leq -2.88\}$?

13. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución de población normal con un valor conocido de σ.

a. Para probar las hipótesis H_0: $\mu = \mu_0$, contra H_a: $\mu > \mu_0$ (donde μ_0 es un número fijo), demuestre que la prueba con el estadístico de prueba $\overline{X}$ y región de rechazo $\bar{x} \geq \mu_0 + 2.33\sigma/\sqrt{n}$ tiene un nivel de significancia de .01.

b. Suponga que se utiliza el procedimiento del inciso (a) para probar H_0: $\mu \leq \mu_0$ contra H_a: $\mu > \mu_0$. Si $\mu_0 = 100$, $n = 25$ y $\sigma = 5$, ¿cuál es la probabilidad de cometer un error de tipo I cuando $\mu = 99$? ¿Cuando $\mu = 98$? En general, ¿qué se puede decir sobre la probabilidad de un error de tipo I cuando el valor real de μ es menor que μ_0? Verifique su aseveración.

14. Reconsidere la situación del ejercicio 11 y suponga que la región de rechazo es $\{\bar{x}: \bar{x} \geq 10.1004$ o $\bar{x} \leq 9.8940\} = \{z: z \geq 2.51$ o $z \leq -2.65\}$.

a. ¿Cuál es α para este procedimiento?

b. ¿Cuál es β cuando $\mu = 10.1$? ¿Cuando $\mu = 9.9$? ¿Es ésta deseable?

8.2 Pruebas sobre una media de población

La discusión general en el capítulo 7 de intervalos de confianza para una media de población μ se enfocó en tres casos diferentes. A continuación se desarrollan procedimientos de prueba para estos casos.

Caso I: Una población normal con σ conocida

Aun cuando la suposición de que el valor de σ es conocido rara vez se cumple en la práctica, este caso proporciona un buen punto de partida debido a la facilidad con la que los procedimientos generales y sus propiedades pueden ser desarrollados. La hipótesis nula en los tres casos propondrá que μ tiene un valor numérico particular, el *valor nulo*, el cual será denotado por μ_0. Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n de la población normal. Entonces la media muestral $\overline{X}$ tiene una distribución normal con valor esperado $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. Cuando H_0 es verdadera $\mu_{\overline{X}} = \mu_0$. Considérese ahora el estadístico Z obtenido estandarizando $\overline{X}$ dada la suposición de que H_0 es verdadera:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Al sustituir la media muestral calculada $\bar{x}$ se obtiene z, la distancia entre $\bar{x}$ y μ_0 expresada en “unidades de desviación estándar”. Por ejemplo, si la hipótesis nula es H_0: $\mu = 100$, $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 10/\sqrt{25} = 2.0$ y $\bar{x} = 103$, entonces el valor estadístico de prueba es $z = (103 - 100)/2.0 = 1.5$. Es decir, el valor observado de $\bar{x}$ es 1.5 desviaciones estándar (de $\overline{X}$) más grande de lo que se espera que sea cuando H_0 es verdadera. El estadístico Z es una medida natural de la distancia entre $\overline{X}$, el estimador de μ, y su valor esperado cuando H_0 es verdadera. Si esta distancia es demasiado grande en una dirección consistente con H_a, la hipótesis nula deberá ser rechazada.

Supóngase primero que la hipótesis alternativa tiene la forma H_a: $\mu > \mu_0$. Entonces un valor de $\bar{x}$ menor que μ_0 indudablemente no apoya a H_a. Tal $\bar{x}$ corresponde a un valor negativo de z (puesto que $\bar{x} - \mu_0$ es negativo y el divisor de $\sigma/\sqrt{n}$ es positivo). Del mismo modo, un valor de $\bar{x}$ que exceda de μ_0 por sólo una pequeña cantidad (correspondiente a z la cual es positiva aunque pequeña) no sugiere que H_0 deberá ser rechazada a favor de H_a. El rechazo de H_0 es apropiado sólo cuando $\bar{x}$ excede considerablemente de μ_0; es decir, cuando el valor de z es positivo y grande. En suma, la región de rechazo apropiada basada en el estadístico de prueba Z en lugar de $\bar{X}$ tiene la forma $z \geq c$.

Como se discutió en la sección 8.1, el valor de corte c deberá ser elegido para controlar la probabilidad de un error de tipo I al nivel α deseado. Esto es fácil de lograr porque la distribución del estadístico de prueba Z cuando H_0 es verdadera es la distribución normal estándar (es por eso que μ_0 se restó al estandarizar). El valor c de corte requerido es el valor crítico z que captura el área de la cola superior α bajo la curva z. Como ejemplo, sea $c = 1.645$, el valor que captura el área de cola .05 ($z_{.05} = 1.645$). Entonces,

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando } H_0 \text{ es verdadera})\\ &= P(Z \geq 1.645 \text{ cuando } Z \sim N(0,1)) = 1 - \Phi(1.645) = .05\end{aligned}$$

Más generalmente, la región de rechazo $z \geq z_\alpha$ tiene una probabilidad α de error de tipo I. El procedimiento de prueba es de *cola superior* porque la región de rechazo se compone sólo de valores grandes del estadístico de prueba.

Un razonamiento análogo para la hipótesis alternativa H_a: $\mu < \mu_0$ sugiere una región de rechazo de la forma $z \leq c$, donde c es un número negativo adecuadamente seleccionado ($\bar{x}$ aparece muy debajo de μ_0 si y sólo si z es bastante negativo). Como Z tiene una distribución normal estándar cuando H_0 es verdadera, con $c = -z_\alpha$ da $P(\text{error de tipo I}) = \alpha$. Ésta es una prueba de *cola inferior*. Por ejemplo, $z_{.10} = 1.28$ implica que la región de rechazo $z \leq -1.28$ especifica una prueba con nivel de significancia de .10.

Por último, cuando la hipótesis alternativa es H_a: $\mu \neq \mu_0$, H_0 deberá ser rechazada si $\bar{x}$ está muy lejos de alguno de los lados de μ_0. Esto equivale a rechazar H_0 si $z \geq c$ o si $z \leq -c$. Supóngase que se desea $\alpha = .05$. Entonces,

$$\begin{aligned}.05 &= P(Z \geq c \text{ o } Z \leq -c \text{ cuando } Z \text{ tiene una distribución normal estándar})\\ &= \Phi(-c) + 1 - \Phi(c) = 2[1 - \Phi(c)]\end{aligned}$$

Por consiguiente c es tal que $1 - \Phi(c)$, el área bajo la curva z a la derecha de c, es .025 (¡y no .05!) De acuerdo con la sección 4.3 o la tabla A.3 del apéndice, $c = 1.96$ y la región de rechazo es $z \geq 1.96$ *o* $z \leq -1.96$. Con cualquier α, la región de rechazo de *dos colas* $z \geq z_{\alpha/2}$ *o* $z \leq -z_{\alpha/2}$ tiene una probabilidad α de error de tipo I (puesto que el área $\alpha/2$ está capturada debajo de cada una de las dos colas de la curva z). De nueva cuenta, la razón clave para utilizar el estadístico de prueba estandarizado Z es que como Z tiene una distribución conocida cuando H_0 es verdadera (normal estándar), es fácil de obtener una región de rechazo con probabilidad de error de tipo I deseada mediante un valor crítico apropiado.

El procedimiento de prueba en el caso I se resume en el cuadro adjunto y las regiones de rechazo correspondientes se ilustran en la figura 8.2.

Hipótesis nula: H_0: $\mu = \mu_0$

Valor del estadístico de prueba: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Hipótesis alternativa	Región de rechazo para la prueba de nivel α
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (prueba de cola superior)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (prueba de cola inferior)
H_a: $\mu \neq \mu_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ (prueba de dos colas)

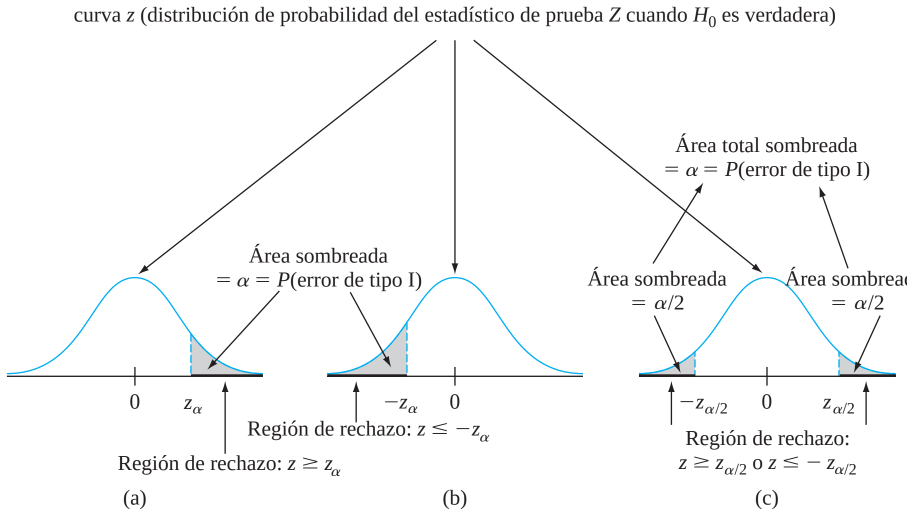


Figura 8.2 Regiones de rechazo para pruebas z: (a) prueba de cola superior; (b) prueba de cola inferior; (c) prueba de dos colas

Se recomienda utilizar la siguiente secuencia de pasos cuando se prueben hipótesis con respecto a un parámetro.

1. Identificar el parámetro de interés y describirlo en el contexto de la situación del problema.
2. Determinar el valor nulo y formular la hipótesis nula.
3. Formular la hipótesis alternativa apropiada

4. Dar la fórmula para el valor calculado del estadístico de prueba (sustituyendo el valor nulo y los valores conocidos de cualesquiera otros parámetros, pero *no* aquellos de cualesquiera cantidades basadas en una muestra).
5. Establecer la región de rechazo para el nivel de significancia seleccionado α.
6. Calcular cualquier cantidad muestral necesaria, sustituir en la fórmula para el valor estadístico de prueba y calcular dicho valor.
7. Decidir si H_0 debe ser rechazada y expresar esta conclusión en el contexto del problema.

La formulación de hipótesis (pasos 2 y 3) deberá ser realizada antes de examinar los datos.

Ejemplo 8.6 Un fabricante de sistemas rociadores utilizados como protección contra incendios en edificios de oficinas afirma que la temperatura de activación del sistema promedio verdadera es de 130°. Una muestra de $n = 9$ sistemas, cuando se somete a prueba, da una temperatura de activación promedio muestral de 131.08°F. Si la distribución de los tiempos de activación es normal con desviación estándar de 1.5°F, ¿contradicen los datos la afirmación del fabricante a un nivel de significancia $\alpha = .01$?

1. Parámetro de interés: μ = temperatura de activación promedio verdadera.
2. Hipótesis nula: H_0: $\mu = 130$ (valor nulo $= \mu_0 = 130$).
3. Hipótesis alternativa: H_a: $\mu \neq 130$ (un alejamiento del valor declarado en *una u otra* dirección es de interés).
4. Valor del estadístico de prueba:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

5. Región de rechazo: la forma de H_a implica el uso de una prueba de dos colas con región de rechazo de $z \geq z_{.005}$ o $z \leq -z_{.005}$. De acuerdo con la sección 4.3 o la tabla A.3 del apéndice, $z_{.005} = 2.58$, así que se rechazaría H_0 si $z \geq 2.58$ o $z \leq -2.58$.

6. Sustituyendo $n = 9$ y $\bar{x} = 131.08$,

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = \frac{1.08}{.5} = 2.16$$

Es decir, la media muestral observada es de un poco más de 2 desviaciones estándar sobre el valor que era de esperarse si H_0 fuera verdadera.

7. El valor calculado $z = 2.16$ no queda en la región de rechazo $(-2.58 < 2.16 < 2.58)$, así que H_0 no puede ser rechazada al nivel de significancia de .01. Los datos no apoyan fuertemente la afirmación de que el promedio verdadero difiere del valor de diseño de 130. ■

β y determinación del tamaño de la muestra Las pruebas z para el caso 1 se encuentran entre las pocas en estadística para las cuales existen fórmulas simples disponibles para β, la probabilidad de un error de tipo II. Considérese en primer lugar la prueba de cola superior con región de rechazo $z \geq z_\alpha$. Ésta equivale a $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$, por lo que H_0 no será rechazada si $\bar{x} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$. Ahora μ' denota un valor particular de μ que excede el valor nulo μ_0. Entonces,

$$\begin{aligned}
\beta(\mu') &= P(H_0 \text{ no es rechazada cuando } \mu = \mu') \\
&= P(\bar{X} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n} \text{ cuando } \mu = \mu') \\
&= P\left(\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu'\right) \\
&= \Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Conforme μ' se incrementa $\mu_0 - \mu'$ se vuelve más negativa, de modo que $\beta(\mu')$ será pequeña cuando μ' excede μ_0 en gran medida (porque el valor con el que se evalúa Φ será entonces bastante negativo). Las probabilidades de error para las pruebas de cola inferior y de dos colas se deducen de manera análoga.

Si σ es grande, la probabilidad de un error de tipo II puede ser grande con un valor alternativo de μ' que sea de interés particular para un investigador. Supóngase que se fija α y que también se especifica β para semejante valor alternativo. En el ejemplo de los sistemas rociadores, los oficiales de la compañía podrían considerar $\mu' = 132$ como un alejamiento muy sustancial de H_0: $\mu = 130$ y desear por consiguiente $\beta(132) = .10$ además de $\alpha = .01$. Más generalmente, considérense las dos restricciones $P(\text{error de tipo I}) = \alpha$ y $\beta(\mu') = \beta$ para α, μ' y β especificadas. Entonces para una prueba de cola superior, el tamaño de la muestra n debe ser elegido para satisfacer

$$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \beta$$

Esto implica que

$$-z_\beta = \begin{array}{l}\text{valor } z \text{ crítico que captura} \\ \text{el área de cola inferior } \beta\end{array} = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}$$

Es fácil resolver esta ecuación para la n deseada. Un argumento paralelo da el tamaño de muestra necesario para las pruebas de cola inferior y de dos colas como se resume en el siguiente recuadro.

Hipótesis alternativa	Probabilidad de error de tipo II $\beta(\mu')$ para una prueba de nivel α
H_a: $\mu > \mu_0$	$\Phi\left(z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

donde $\Phi(z) =$ función de distribución acumulativa normal estándar.

El tamaño de la muestra n con el cual una prueba de nivel α también tiene $\beta(\mu') = \beta$ con el valor alternativo μ' es

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{prueba de una cola (superior o inferior)} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{para una prueba de dos colas (una solución aproximada)} \end{cases}$$

Ejemplo 8.7 Sea μ la vida promedio verdadera de la banda de rodamiento de un cierto tipo de neumático. Considere poner a prueba H_0: $\mu = 30{,}000$ contra H_a: $\mu > 30{,}000$ basado en un tamaño de muestra $n = 16$ de una distribución de población normal con $\sigma = 1500$. Una prueba con $\alpha = .01$ requiere $z_\alpha = z_{.01} = 2.33$. La probabilidad de cometer un error de tipo II cuando $\mu = 31{,}000$ es

$$\beta(31{,}000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Como $z_{.1} = 1.28$, el requerimiento de que el nivel de prueba .01 también tenga $\beta(31{,}000) = .1$ necesita

$$n = \left[\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}\right]^2 = (-5.42)^2 = 29.32$$

El tamaño de muestra debe ser un entero, por lo tanto se deberán utilizar 30 neumáticos. ■

Caso II: Pruebas con muestras grandes

Cuando el tamaño de muestra es grande, las pruebas z en el caso I son fáciles de modificar para dar procedimientos de prueba válidos sin requerir una distribución de población normal o σ conocida. El resultado clave se utilizó en el capítulo 7 para justificar intervalos de confianza para muestra grande: una n grande implica que la variable estandarizada

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene *aproximadamente* una distribución normal estándar. La sustitución del valor nulo μ_0 en lugar de μ da el estadístico de prueba

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

que tiene de manera aproximada una distribución normal estándar cuando H_0 es verdadera. El uso de las regiones de rechazo dadas previamente para el caso I (p. ej., $z \geq z_\alpha$ cuando la hipótesis alternativa es H_a: $\mu > \mu_0$) resulta entonces en procedimientos de prueba con los cuales el nivel de significancia es casi (y no exactamente) α. Se utilizará de nuevo la regla empírica $n > 40$ para caracterizar un tamaño de muestra grande.

Ejemplo 8.8 Se utiliza un penetrómetro cónico dinámico para medir la resistencia de un material a la penetración (mm/golpe), a medida que el cono es insertado en pavimento o subsuelo. Suponga que para una aplicación particular, se requiere que el valor de penetración cónica dinámica promedio verdadera para un cierto tipo de pavimento sea menor que 30. El pavimento no será utilizado a menos que exista evidencia concluyente de que la especificación se satisfizo. Formule y pruebe las hipótesis apropiadas utilizando los datos siguientes (“Probabilistic Model for the Analysis of Dynamic Cone Penetrometer Test Values in Pavement Structure Evaluation”, *J. of Testing and Evaluation*, 1999: 7–14:

14.1	14.5	15.5	16.0	16.0	16.7	16.9	17.1	17.5	17.8
17.8	18.1	18.2	18.3	18.3	19.0	19.2	19.4	20.0	20.0
20.8	20.8	21.0	21.5	23.5	27.5	27.5	28.0	28.3	30.0
30.0	31.6	31.7	31.7	32.5	33.5	33.9	35.0	35.0	35.0
36.7	40.0	40.0	41.3	41.7	47.5	50.0	51.0	51.8	54.4
55.0	57.0								

La figura 8.3 muestra un resumen descriptivo obtenido con Minitab. La penetración cónica dinámica media muestral es menor que 30. Sin embargo, existe una cantidad sustancial de variación en los datos (coeficiente de variación muestral $= s/\bar{x} = .4265$), de modo que el hecho de que la media sea menor que el valor de corte de la especificación de diseño puede ser simplemente una consecuencia de la variabilidad muestral. Obsérvese que el histograma no se asemeja en absoluto a una curva normal (y una gráfica de probabilidad normal no exhibe un patrón lineal), aunque las pruebas z con muestras grandes no requieren una distribución de población normal.

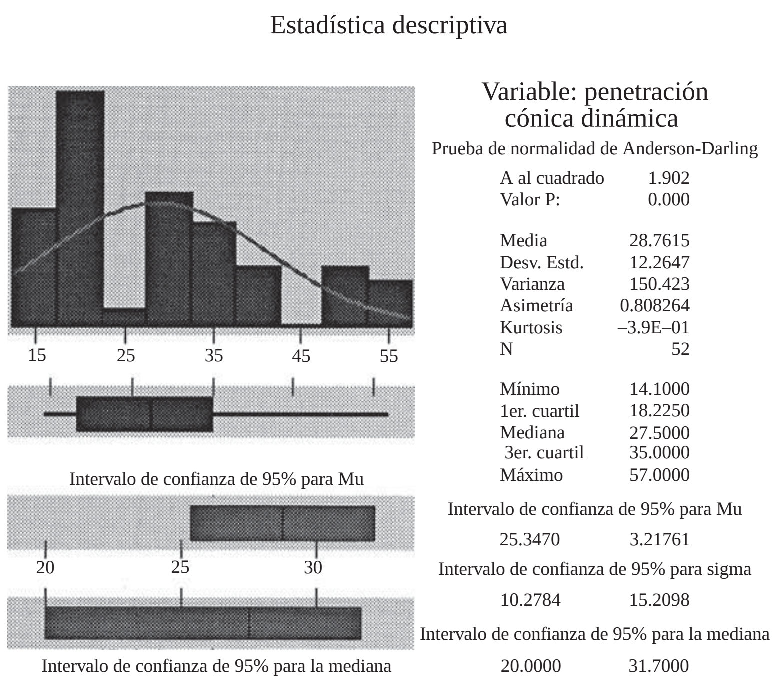


Figura 8.3 Resumen descriptivo generado por Minitab para los datos de penetración cónica dinámica del ejemplo 8.8

1. μ = valor de penetración cónica dinámica promedio verdadero
2. H_0: $\mu = 30$

3. H_a: $\mu < 30$ (por consiguiente el pavimento no será utilizado a menos que la hipótesis nula sea rechazada)

4. $z = \dfrac{\bar{x} - 30}{s/\sqrt{n}}$

5. Una prueba con nivel de significancia de .05 rechaza a H_0 cuando $z \leq -1.645$ (una prueba de cola inferior).

6. Con $n = 52$, $\bar{x} = 28.76$ y $s = 12.2647$,

$$z = \frac{28.76 - 30}{12.2647/\sqrt{52}} = \frac{-1.24}{1.701} = -.73$$

7. Como $-.73 > -1.645$, H_0 no puede ser rechazada. No se cuenta con evidencia precisa para concluir que $\mu < 30$; el uso del pavimento no se justifica. ■

La determinación de β y el tamaño de muestra necesario para estas pruebas con muestra grande pueden basarse en la especificación de un valor adecuado de σ y en el uso de las fórmulas para el caso I (aun cuando se utilice s en la prueba) o en el uso de la metodología que se introducirá en breve en conexión con el caso III.

Caso III: Una distribución de población normal

Cuando n es pequeño, el teorema del límite central TLC ya no puede ser invocado para justificar el uso de una prueba con muestra grande. Esta misma dificultad se presentó al obtener un intervalo de confianza (IC) con muestra pequeña para μ en el capítulo 7. El método utilizado aquí es el mismo que se empleó allí: se supondrá que la distribución de población es por lo menos aproximadamente normal y se describirán los procedimientos de prueba cuya validez se fundamenta en esta suposición. Si un investigador tiene una buena razón para creer que la distribución de población es bastante no normal, se puede utilizar una prueba libre de distribución del capítulo 15. Alternativamente, un estadístico puede ser consultado en cuanto a procedimientos válidos para familias específicas de distribuciones de población aparte de la familia normal. O se puede desarrollar un procedimiento bootstrap.

En el capítulo 7 se utilizó el resultado clave en el cual están basadas las pruebas con una media de población normal para obtener el intervalo de confianza t para una muestra: si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución normal, la variable estandarizada

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución t con $n - 1$ grados de libertad (gl). Considérese poner a prueba H_0: $\mu = \mu_0$ contra H_a: $\mu > \mu_0$ utilizando el estadístico de prueba $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$. Es decir, el estadístico de prueba resulta de estandarizar $\bar{X}$ conforme a la suposición de que H_0 es verdadera (utilizando $S/\sqrt{n}$, la desviación estándar estimada de $\bar{X}$, en lugar de $\sigma/\sqrt{n}$). Cuando H_0 es verdadera, el estadístico de prueba tiene una distribución t con $n - 1$ grados de libertad. El conocimiento de la distribución del estadístico de prueba cuando H_0 es verdadera (la "distribución nula") permite construir una región de rechazo para la cual la probabilidad de error de tipo I se controla al nivel deseado. En particular, el uso del valor crítico t de cola superior $t_{\alpha,n-1}$ para especificar la región de rechazo $t \geq t_{\alpha,n-1}$ implica que

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(T \geq t_{\alpha,n-1} \text{ cuando } T \text{ tiene una distribución } t \text{ con } n - 1 \\ &\quad \text{grados de libertad}) \\ &= \alpha \end{aligned}$$

El estadístico de prueba es en realidad el mismo del caso de muestra grande pero se designa T para recalcar que su distribución nula es una distribución t con $n - 1$ grados de libertad en lugar de la distribución normal estándar (z). La región de rechazo para la prueba t difiere de aquella para la prueba z sólo en que un valor crítico $t_{\alpha,n-1}$ reemplaza al valor crítico z_α de z. Comentarios similares se aplican a alternativas para las cuales una prueba de cola inferior o de dos colas es apropiada.

Prueba *t* con una muestra

Hipótesis nula: H_0: $\mu = \mu_0$

Valor estadístico de prueba: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Hipótesis alternativa	Región de rechazo para una prueba de nivel α
H_a: $\mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (cola superior)
H_a: $\mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (cola inferior)
H_a: $\mu \neq \mu_0$	$t \geq t_{\alpha/2,n-1}$ o $t \leq -t_{\alpha/2,n-1}$ (dos colas)

Ejemplo 8.9 El glicerol es un importante subproducto de la fermentación de etanol en la producción de vino y contribuye a la dulzura, cuerpo y la plenitud de los vinos. El artículo "A Rapid and Simple Method for Simultaneous Determination of Glycerol, Fructose, and Glucose in Wine" (*American J. of Enology and Viticulture*, 2007: 279−283) incluye las siguientes observaciones sobre la concentración de glicerol (mg/ml) para muestras de vinos blancos de calidad estándar (sin certificar): 2.67, 4.62, 4.14, 3.81, 3.83. Supongamos que el valor de la concentración deseada es de 4. ¿Los datos de la muestra indican que la concentración promedio real es algo más que el valor deseado? El siguiente diagrama de probabilidad normal de Minitab proporciona un fuerte apoyo para suponer que la distribución de la población de la concentración de glicerol es normal. Se llevará a cabo una prueba de hipótesis adecuada utilizando la prueba t de una muestra con un nivel de significancia de .05.

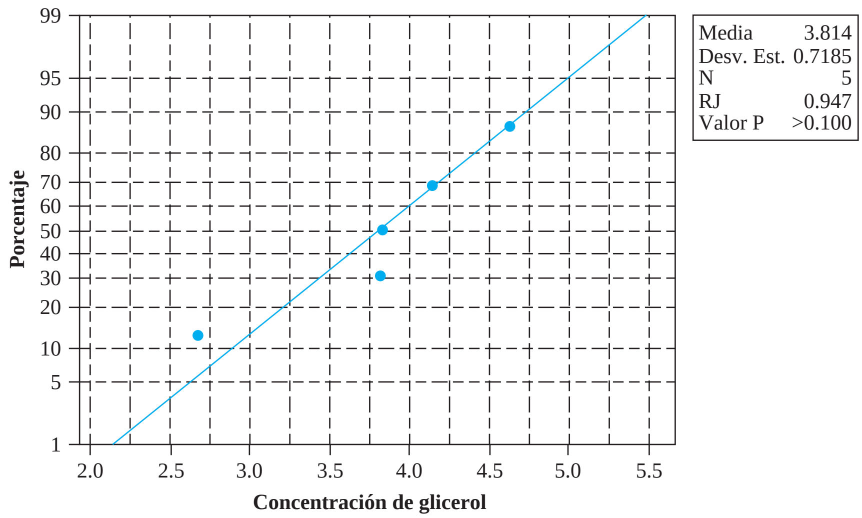


Figura 8.4 Gráfica de probabilidad normal para los datos del ejemplo 8.9

1. $\mu =$ promedio real de la concentración de glicerol

2. H_0: $\mu = 4$

3. H_a: $\mu \neq 4$

4. $t = \dfrac{\bar{x} - 4}{s/\sqrt{n}}$

5. La desigualdad en H_a implica que una prueba de dos colas es apropiada, lo que requiere $t_{\alpha/2,n-1} = t_{025,4} = 2.776$. Por lo tanto H_0 se rechaza si $t \geq 2.776$ o $t \leq -2.776$.

6. $\sum x_i = 19.07$ y $\sum x_i^2 = 74.7979$ con las cuales $\bar{x} = 3.814$, $s = .718$ y el error estándar estimado de la media es $s/\sqrt{n} = .321$. El valor del estadístico de prueba es entonces $t = (3.814 - 4)/.321 = -.58$.

7. Es evidente que $t = -.58$ no se encuentra en la región de rechazo para un nivel de significancia de .05. Todavía es posible que $\mu = 4$. La desviación de la media de la muestra 3.814 de su valor esperado 4 cuando H_0 es cierta se puede atribuir simplemente a la variabilidad del muestreo en lugar de que H_0 sea falsa.

Los datos de salida de Minitab resultan de una solicitud para realizar una prueba t de dos colas de una muestra y presenta valores calculados idénticos a los que se acaban de obtener. El hecho de que el último número en la salida, el "valor P" sea superior a .05 (y cualquier otro nivel de significancia razonable) implica que la hipótesis nula no puede rechazarse. Esto se analiza con detalle en la sección 8.4.

```
Test of mu = 4 vs not = 4
Variable   N   Mean  StDev  SE Mean  95% CI              T      P
glyc conc  5  3.814  0.718    0.321  (2.922, 4.706)  −0.58  0.594
```

***β* y determinación del tamaño de la muestra** El cálculo de β con el valor alternativo μ' en el caso I se realizó expresando la región de rechazo en función de $\bar{x}$ (p. ej., $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$) y luego restando μ' para estandarizar correctamente. Un método equivalente implica observar que cuando $\mu = \mu'$, el estadístico de prueba $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$ sigue teniendo una distribución normal con varianza 1, pero ahora el valor medio de Z está dado por $(\mu' - \mu_0)/(\sigma/\sqrt{n})$. Es decir, cuando $\mu = \mu'$, el estadístico de prueba sigue teniendo una distribución normal pero no la distribución normal estándar. Por eso, $\beta(\mu')$ es un área bajo la curva normal correspondiente al valor medio $(\mu' - \mu_0)/(\sigma/\sqrt{n})$ y varianza 1. Tanto α como β implican trabajar con variables normalmente distribuidas.

El cálculo de $\beta(\mu')$ para la prueba t es mucho menos directo. Esto es porque la distribución del estadístico de prueba $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ es bastante complicada cuando H_0 es falsa y H_a es verdadera. Por consiguiente, en una prueba de cola superior, determinar

$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ cuando } \mu = \mu' \text{ en lugar de } \mu_0)$$

implica integrar una desagradable función de densidad. Esto debe hacerse numéricamente. Los resultados se resumen en gráficas de β que aparecen en la tabla A.17 del apéndice. Existen cuatro juegos de gráficas, correspondientes a pruebas de una cola a nivel .05 y nivel .01 y pruebas de dos colas a los mismos niveles.

Para entender cómo se utilizan estas gráficas, obsérvese primero que tanto β como el tamaño de muestra necesario n en el caso I son funciones no sólo de la diferencia absoluta $|\mu_0 - \mu'|$ sino de $d = |\mu_0 - \mu'|/\sigma$. Supóngase, por ejemplo, que $|\mu_0 - \mu'| = 10$. Este alejamiento de H_0 será mucho más fácil de descubrir (β más pequeña) cuando $\sigma = 2$, en cuyo caso μ_0 y μ' están a 5 desviaciones estándar de la población una de otra, que cuando $\sigma = 10$. El hecho de que β para la prueba t dependa de d y no sólo de $|\mu_0 - \mu'|$ es desafortunado, puesto que para utilizar las gráficas se debe tener alguna idea del valor verdadero de σ. Una suposición conservadora (grande) para σ dará por resultado un valor conservador (grande) de $\beta(\mu')$ y una estimación conservadora del tamaño de muestra necesario para α y $\beta(\mu')$ prescritas.

Una vez que se seleccionan la μ' alternativa y el valor de σ, se calcula d y su valor se localiza sobre el eje horizontal del conjunto de curvas pertinente. El valor de β es la altura de la curva con $n - 1$ grados de libertad por encima del valor de d (es necesaria una interpolación visual si $n - 1$ no es un valor para el cual la curva correspondiente aparece), como se ilustra en la figura 8.5.

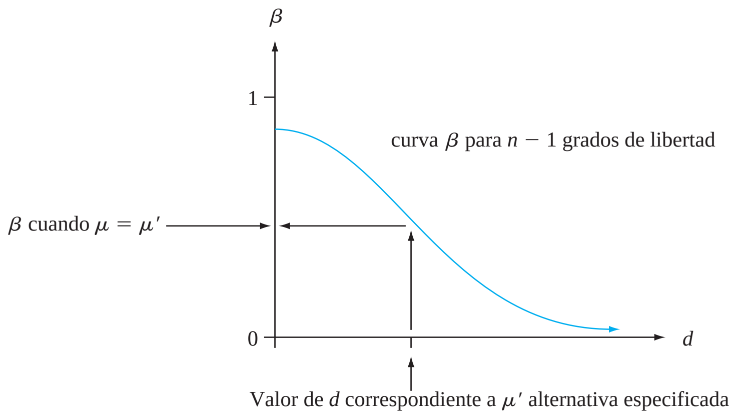


Figura 8.5 Curva β típica de la prueba t

En lugar de fijar n (es decir, $n - 1$) y por consiguiente la curva particular en donde se lee β) se podría prescribir tanto α (.05 o .01 en este caso) y un valor de β para las μ' y σ seleccionadas. Después de calcular d, se localiza el punto (d, β) en el conjunto de gráficas pertinentes. La curva debajo y más próxima a este punto da $n - 1$ y por consiguiente n (de nuevo con frecuencia se requiere interpolación).

Ejemplo 8.10 Se supone que la caída de voltaje promedio verdadera entre el colector y el emisor de transistores bipolares de compuerta aislados de cierto tipo es cuando mucho de 2.5 volts. Un investigador selecciona una muestra de $n = 10$ de esos transistores y utiliza los voltajes resultantes para probar H_0: $\mu = 2.5$ contra H_a: $\mu > 2.5$ por medio de una prueba t con nivel de significancia $\alpha = .05$. Si la desviación estándar de la distribución de voltaje es $\sigma = .100$, ¿qué tan probable es que H_0 no será rechazada cuando en realidad $\mu = 2.6$? Con $d = |2.5 - 2.6|/.100 = 1.0$, el punto sobre la curva β con 9 grados de libertad para una prueba de una cola con $\alpha = .05$ por encima de 1.0 tiene aproximadamente .1 de altura, por lo tanto $\beta \approx .1$. El investigador podría pensar que éste es un valor de β demasiado grande con semejante alejamiento sustancial de H_0 y puede que desee tener $\beta = .05$ con este valor alternativo de μ. Como $d = 1.0$, el punto $(d, \beta) = (1.0, .05)$ debe ser localizado. Este punto se aproxima mucho a la curva de 14 grados de libertad, por lo tanto con $n = 15$ se obtendrá tanto $\alpha = .05$ como $\beta = .05$ cuando el valor de μ es 2.6 y $\sigma = .10$. Un valor más grande de σ daría una β más grande para esta alternativa y un valor alternativo de μ más cercano a 2.5 también daría por resultado un valor incrementado de β. ■

La mayoría de los programas de computadora estadísticos también calcularán probabilidades de error de tipo II. Por lo general, trabajan en términos de **potencia**, que es simplemente $1 - \beta$. Un valor pequeño de β (cerca de 0) es equivalente a una potencia grande (cerca de 1). Una prueba de *gran alcance* es la que tiene gran potencia y por tanto buena capacidad para detectar cuándo la hipótesis nula es falsa.

Como ejemplo, se le pide a Minitab que determine la potencia de la prueba de cola superior del ejemplo 8.10 para tres tamaños de muestra de 5, 10 y 15 cuando $\alpha = .05$, $\sigma = .10$ y el valor de μ es en realidad 2.6 en vez del valor nulo de 2.5, una "diferencia" de

2.6 − 2.5 = .1. También se le pidió al software determinar el tamaño de la muestra necesario para una potencia de .9 (β = .1) y .95. Los datos de salida se dan a continuación.

```
Power and Sample Size
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.1
                Sample
Difference       Size    Power
0.1               5      0.579737
0.1              10      0.897517
0.1              15      0.978916
                   Sample Target
                                           Actual
Difference       Size    Power             Power
0.1              11      0.90              0.924489
0.1              13      0.95              0.959703
```

La potencia para el tamaño de muestra $n = 10$ es un poco menor que .9. Así que si se insiste en que la potencia sea de por lo menos .9, es necesaria una muestra de tamaño 11 y la potencia real para n es aproximadamente .92. El software dice que para una potencia objetivo de .95, es necesario un tamaño de muestra de $n = 13$, mientras que echando un vistazo a nuestras curvas β dio 15. Cuando está disponible, este tipo de software es más confiable que las curvas. Por último, Minitab ahora también proporciona curvas de potencia para los tamaños de muestra determinados, tal como se muestra en la figura 8.6. Estas curvas muestran cómo aumenta la potencia para cada tamaño de muestra a medida que el valor real de μ se desplaza más allá y más lejos del valor nulo.

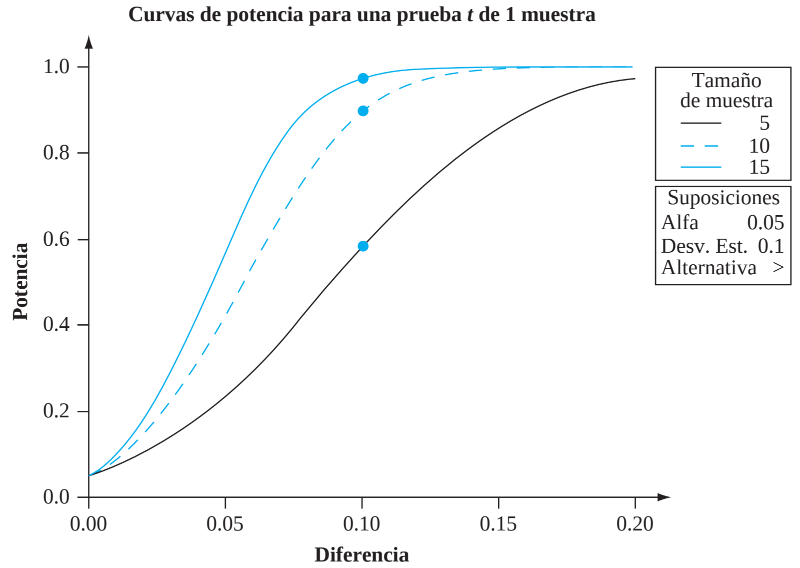


Figura 8.6 Curvas de potencia de Minitab para la prueba t del ejemplo 8.10

EJERCICIOS Sección 8.2 (15–36)

15. Sea un estadístico de prueba Z que tiene una distribución normal estándar cuando H_0 es verdadera. Dé el nivel de significancia en cada una de las siguientes situaciones:

a. H_a: $\mu > \mu_0$, región de rechazo $z \geq 1.88$
b. H_a: $\mu < \mu_0$, región de rechazo $z \leq -2.75$
c. H_a: $\mu \neq \mu_0$, región de rechazo $z \geq 2.88$ o $z \leq -2.88$

16. Sea el estadístico de prueba T que tiene una distribución t cuando H_0 es verdadera. Dé el nivel de significancia en cada una de las situaciones:

a. H_a: $\mu > \mu_0$, grados de libertad $= 15$, región de rechazo $t \geq 3.733$

b. H_a: $\mu < \mu_0$, $n = 24$, región de rechazo $t \leq -2.500$

c. H_a: $\mu \neq \mu_0$, $n = 31$, región de rechazo $t \geq 1.697$ o $t \leq -1.697$

17. Responda las siguientes preguntas en relación con el problema de los neumáticos en el ejemplo 8.7.

a. Si $\bar{x} = 30{,}960$ y se utiliza una prueba de nivel $\alpha = .01$, ¿cuál es la decisión?

b. Si utiliza una prueba de nivel .01, ¿cuál es $\beta(30{,}500)$?

c. Si se utiliza una prueba de nivel .01 y también se requiere que $\beta(30{,}500) = .05$, ¿qué tamaño de muestra n es necesario?

d. Si $\bar{x} = 30{,}960$ ¿cuál es la α más pequeña con la cual H_0 puede ser rechazada (con base en $n = 16$)?

18. Reconsidere la situación de secado de pintura del ejemplo 8.2, en el cual el tiempo de secado para un espécimen de prueba está normalmente distribuido con $\sigma = 9$. Las hipótesis H_0: $\mu = 75$ contra H_a: $\mu < 75$ tienen que ser probadas con una muestra aleatoria de $n = 25$ observaciones.

a. ¿A cuántas desviaciones estándar (de $\bar{X}$) por debajo del valor nulo se encuentra $\bar{x} = 72.3$?

b. Si $\bar{x} = 72.3$, ¿cuál es la conclusión si utiliza $\alpha = .01$?

c. ¿Cuál es α para el procedimiento de prueba que rechaza H_0 cuando $z \leq -2.88$?

d. Con el procedimiento de prueba del inciso (c), ¿cuál es $\beta(70)$?

e. Si se utiliza el procedimiento de prueba del inciso (c), ¿qué n es necesaria para garantizar $\beta(70) = .01$?

f. Si se utiliza un prueba de nivel .01 con $n = 100$, ¿cuál es la probabilidad de un error de tipo I cuando $\mu = 76$?

19. Se determinó el punto de fusión de cada una de las 16 muestras de una marca de aceite vegetal hidrogenado y el resultado fue $\bar{x} = 94.32$. Suponiendo que la distribución del punto de fusión es normal con $\sigma = 1.20$.

a. Probar H_0: $\mu = 95$ contra H_a: $\mu \neq 95$ por medio de una prueba de dos colas de nivel .01.

b. Si se utiliza una prueba de nivel .01, ¿cuál es $\beta(94)$, la probabilidad de un error de tipo II cuando $\mu = 94$?

c. ¿Qué valor de n es necesario para garantizar que $\beta(94) = .1$ cuando $\alpha = .01$?

20. Se anuncia que focos de un tipo duran un promedio de 750 horas. El precio de estos focos es muy favorable por lo que un cliente potencial ha decidido continuar con un convenio de compra hasta que concluyentemente se demuestre que la duración promedio verdadera sea menor que la anunciada. Se seleccionó una muestra aleatoria de 50 focos, se determinó la duración de cada uno, se probaron las hipótesis apropiadas con Minitab y se obtuvieron los siguientes resultados.

```
Variable   N    Mean  StDev  SEMean      Z  P-Value
lifetime  50  738.44  38.20    5.40  -2.14    0.016
```

¿Qué conclusión sería apropiada para un nivel de significancia de .05? ¿Un nivel de significancia de .01? ¿Qué nivel de significancia y conclusión recomendaría?

21. Se supone que el diámetro promedio verdadero de cojinetes de bolas de un tipo es de .5 pulg. Se realizará una prueba t con una muestra para ver si éste es el caso. ¿Qué conclusión es apropiada en cada una de las siguientes situaciones?

a. $n = 13$, $t = 1.6$, $\alpha = .05$

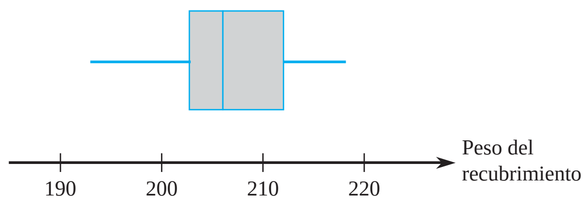
b. $n = 13$, $t = -1.6$, $\alpha = .05$

c. $n = 25$, $t = -2.6$, $\alpha = .01$

d. $n = 25$, $t = -3.9$

22. El artículo "The Foreman's View of Quality Control" (*Quality Engr.*, 1990: 257–280) describe una investigación de pesos de recubrimiento de grandes tuberías resultantes de un proceso de galvanizado. Los estándares de producción demandan un peso promedio verdadero de 200 lb por tubería. El resumen y la gráfica de caja descriptivos adjuntos fueron producidos por Minitab.

```
Variable N   Mean  Median TrMean StDev  SEMean
ctg wt   30 206.73 206.00 206.81  6.35   1.16
Variable    Min     Max     Q1      Q3
ctg wt      193.00  218.00  202.75  212.00
```



a. ¿Qué sugiere la gráfica de caja sobre el estado de la especificación de peso de recubrimiento promedio verdadero?

b. Una gráfica de probabilidad normal de los datos resultó bastante recta. Use los datos de salida descriptivos para probar las hipótesis apropiadas.

23. El ejercicio 36 del capítulo 1 dio $n = 26$ observaciones sobre el tiempo de escape (s) de trabajadores petroleros en un ejercicio simulado, con media y desviación estándar muestrales de 370.69 y 24.36, respectivamente. Suponga que los investigadores creyeron de antemano que el tiempo de escape promedio verdadero sería cuando mucho de 6 min. ¿Contradicen los datos esta creencia anticipada? Suponiendo normalidad, pruebe las hipótesis apropiadas con un nivel de significancia de .05.

24. Reconsidere las observaciones muestrales sobre viscosidad estabilizada de especímenes de asfalto introducidos en el ejercicio 46 del capítulo 1 (2781, 2900, 3013, 2856 y 2888). Suponga que para una aplicación particular se requiere que la viscosidad promedio verdadera sea de 3000. ¿Parece haber sido satisfecho este requerimiento? Formule y pruebe las hipótesis apropiadas.

25. El porcentaje deseado de SiO_2 en cierto tipo de cemento aluminoso es de 5.5. Para comprobar si el porcentaje promedio verdadero es de 5.5 en una instalación de producción particular, se analizaron 16 muestras obtenidas de manera independiente. Suponga que el porcentaje de SiO_2 en una muestra está normalmente distribuido con $\sigma = .3$ y que $\bar{x} = 5.25$.

a. ¿Indica esto concluyentemente que el porcentaje promedio verdadero difiere de 5.5? Realice el análisis siguiendo la secuencia de pasos sugerida en el texto.

b. Si el porcentaje promedio verdadero es $\mu = 5.6$ y se utiliza una prueba de nivel $\alpha = .01$ con $n = 16$, ¿cuál es la probabilidad de descubrir este alejamiento de H_0?

c. ¿Qué valor de n se requiere para satisfacer $\alpha = .01$ y $\beta(5.6) = .01$?

26. Para obtener información sobre las propiedades de resistencia a la corrosión de un tipo de conducto de acero, se enterraron 45 especímenes en el suelo durante 2 años. Se midió entonces la penetración máxima (en mils) en cada espécimen y se obtuvo una penetración promedio muestral de $\bar{x} = 52.7$ y una desviación estándar muestral de $s = 4.8$. Los conductos se fabricaron con la especificación de que la penetración promedio verdadera sea cuando mucho de 50 mils. Se utilizarán a menos que se pueda demostrar concluyentemente que la especificación no ha sido satisfecha. ¿Qué concluiría?

27. La identificación automática de los límites de estructuras significativas en una imagen médica es un área de investigación continua. El artículo "Automatic Segmentation of Medical Images Using Image Registration: Diagnostic and Simulation Applications" (*J. of Medical Engr. and Tech.*, 2005: 53–63) discutió una nueva técnica para realizar la identificación mencionada. Una medida de la precisión de la región automática es el desplazamiento lineal promedio. El artículo dio las siguientes observaciones de desplazamiento lineal promedio con una muestra de 49 riñones (unidades de dimensiones en pixeles).

1.38	0.44	1.09	0.75	0.66	1.28	0.51
0.39	0.70	0.46	0.54	0.83	0.58	0.64
1.30	0.57	0.43	0.62	1.00	1.05	0.82
1.10	0.65	0.99	0.56	0.56	0.64	0.45
0.82	1.06	0.41	0.58	0.66	0.54	0.83
0.59	0.51	1.04	0.85	0.45	0.52	0.58
1.11	0.34	1.25	0.38	1.44	1.28	0.51

a. Resuma y describa los datos.
b. ¿Es factible que el desplazamiento lineal promedio esté por lo menos normalmente distribuido en forma aproximada? ¿Se debe suponer normalidad antes de calcular un intervalo de confianza para el desplazamiento lineal promedio verdadero o probar las hipótesis en cuanto a desplazamiento lineal promedio verdadero? Explique.
c. Los autores comentaron que en la mayoría de los casos el desplazamiento lineal promedio es del orden de 1.0 o mejor. ¿Proporcionan en realidad los datos una fuerte evidencia para concluir que el desplazamiento lineal promedio en estas circunstancias es menor que 1.0? Efectúe una prueba apropiada de hipótesis.
d. Calcule un límite de confianza superior para el desplazamiento lineal promedio verdadero utilizando un nivel de confianza de 95% e interprete este límite.

28. La cirugía menor de caballos en condiciones de campo requiere un anestésico de corta duración confiable que produzca una buena relajación muscular, cambios cardiovasculares y respiratorios mínimos y una rápida y tranquila recuperación con mínimos efectos secundarios de modo que los caballos puedan ser dejados sin atención. El artículo "A Field Trial of Ketamine Anesthesia in the Horse" (*Equine Vet. J.*, 1984: 176–179) reporta que con una muestra de $n = 73$ caballos a los cuales se les administró ketamina en ciertas condiciones, el tiempo de reclinación lateral (echado) promedio muestral fue de 18.86 min y la desviación estándar de 8.6 min. ¿Sugieren estos datos que el tiempo de reclinación lateral promedio verdadera en estas condiciones es menor que 20 min? Pruebe las hipótesis apropiadas a un nivel de significancia de .10.

29. El artículo "Uncertainty Estimation in Railway Track Life-Cycle Cost" (*J. of Rail and Rapid Transit*, 2009) presenta los siguientes datos sobre el tiempo de reparación (minutos) de la rotura de un carril alto en una vía curva del tren de cierta línea de ferrocarril.

159 120 480 149 270 547 340 43 228 202 240 218

Una gráfica de probabilidad normal de los datos muestra un patrón bastante lineal, por lo que es factible que la distribución de la población del tiempo de reparación sea por lo menos aproximadamente normal. La desviación media y estándar de la muestra son 249.7 y 145.1, respectivamente.
a. ¿Hay pruebas de peso para concluir que de verdad el tiempo medio de reparación sea superior a 200 minutos? Lleve a cabo una prueba de hipótesis con un nivel de significancia de .05.
b. Usando $\sigma = 150$, ¿cuál es la probabilidad de error tipo II de la prueba utilizada en el inciso (a) cuando el tiempo promedio de reparación verdadero es en realidad 300 minutos? Es decir, ¿cuál es $\beta(300)$?

30. ¿Alguna vez se ha visto frustrado porque no puede conseguir un contenedor de algún tipo del que se pueda liberar la última parte de su contenido? El artículo "Shake, Rattle, and Squeeze: How Much Is Left in That Container?" (*Consumer Reports*, May 2009: 8), informó sobre una investigación de este tema para varios productos de consumo. Supongamos que cinco tubos de 6.0 onzas de pasta de dientes de una marca en particular son seleccionados al azar y se les aprieta hasta que no haya más pasta de dientes que salga. Luego, se corta cada tubo y la cantidad restante se pesa, dando lugar a los siguientes datos (en consistencia con lo que el citado artículo informaba): .53, .65, .46, .50, .37. ¿Parece que la cantidad promedio restante real es inferior al 10% del contenido neto anunciado?
a. Compruebe la validez de los supuestos necesarios para probar la hipótesis apropiada.
b. Lleve a cabo una prueba de las hipótesis adecuadas con un nivel de significancia de .05. ¿Cambiaría su conclusión si se ha utilizado un nivel de significancia de .01?
c. Describa en contexto los tipos de errores I y II, y diga qué error podría haberse cometido para llegar a una conclusión.

31. Un lugar de trabajo seguro y bien diseñado puede contribuir en gran medida al aumento de la productividad. Es especialmente importante que los trabajadores no realicen tareas que excedan sus capacidades, tales como cargar. Los siguientes datos sobre el peso máximo de carga (PMC, en kg) para una frecuencia de cuatro cargas/min se presentaron en el artículo "The Effects of Speed, Frequency, and Load on Measured Hand Forces for a Floor-to-Knuckle Lifting Task" (*Ergonomics*, 1992: 833−843), los sujetos fueron seleccionados al azar de una población de hombres sanos con edades de 18 a 30 años. Suponiendo que el PMC se distribuye normalmente, ¿los datos sugieren que la media poblacional del PMC supera los 25? Lleve a cabo una prueba con un nivel de significancia de .05.

25.8 36.6 26.3 21.8 27.2

32. La cantidad diaria recomendada de zinc en la dieta entre los varones mayores de 50 años de edad es de 15 mg/día. El artículo "Nutrient Intakes and Dietary Patterns of Older Americans: A National Study" (*J. of Gerontology*, 1992: M145−150) presenta el siguiente resumen de los datos sobre el consumo de zinc en una muestra de varones con edades entre 65–74 años: $n = 115$, $\bar{x} = 11.3$, y $s = 6.43$. ¿Estos datos indican que la ingesta de zinc diaria promedio en la población de hombres de todas las edades de 65 a 74 años cae por debajo de la cantidad recomendada?

33. Reconsidere los datos del ejemplo que muestran la proporción de gastos (%) de los fondos de crecimiento de gran capitalización mutua presentados por primera vez en el ejercicio 1.53.

```
0.52 1.06 1.26 2.17 1.55 0.99 1.10 1.07 1.81 2.05
0.91 0.79 1.39 0.62 1.52 1.02 1.10 1.78 1.01 1.15
```

Una gráfica de probabilidad normal muestra un patrón bastante lineal.

a. ¿Hay pruebas de peso para concluir que la media poblacional de la proporción de gastos excede el 1%? Lleve a cabo una prueba de las hipótesis pertinentes con un nivel de significancia de .01.

b. Volviendo al inciso (a), describa en contexto el tipo de errores I y II y diga qué error podría haberse cometido para llegar a su conclusión. La fuente de la cual se obtuvieron los datos informó que $\mu = 1.33$ para la población de todos los 762 fondos. Así que, ¿ha cometido un error al llegar a su conclusión?

c. Suponiendo que $\sigma = .5$, determine e interprete la potencia de la prueba en el inciso (a) para el valor real de μ indicado en el inciso (b).

34. Una muestra de 12 detectores de radón de un cierto tipo fue seleccionada y cada uno fue expuesto a 100 pCi/L de radón. Las lecturas resultantes son las siguientes:

105.6	90.9	91.2	96.9	96.5	91.3
100.1	105.0	99.6	107.7	103.3	92.4

a. ¿Estos datos sugieren que la lectura media de la población en estas condiciones difiere de 100? Establezca y ponga a prueba las hipótesis adecuadas con $\alpha = .05$.

b. Suponga que antes del experimento se había supuesto un valor de $\sigma = 7.5$. ¿Cuántas determinaciones habrían sido apropiadas entonces para obtener $\beta = .10$ para la $\mu = 95$ alternativa?

35. Demuestre que para cualquier $\Delta > 0$, cuando la distribución de la población es normal y se conoce σ, la prueba de dos colas satisface $\beta(\mu_0 - \Delta) = \beta(\mu_0 + \Delta)$, de modo que $\beta(\mu')$ es simétrica respecto a μ_0.

36. Para un valor μ' alternativo fijo, demuestre que $\beta(\mu') \to 0$ cuando $n \to \infty$, para una prueba z de una cola o de dos colas en el caso de una distribución normal de la población con σ conocida.

8.3 Pruebas relacionadas con una proporción de población

Sea p la proporción de individuos u objetos en una población que poseen una propiedad especial (p. ej., carros con transmisión manual o fumadores que fuman cigarrillos con filtro). Si un individuo u objeto con la propiedad es etiquetado como éxito (S), entonces p es la proporción de éxitos de la población. Las pruebas relacionadas con p se basarán en una muestra aleatoria de tamaño n de la población. Siempre que n sea pequeña con respecto al tamaño de la población, X (el número de éxitos en la muestra) tiene (aproximadamente) una distribución binomial. Además, si n es grande [$np \geq 10$ y $n(1 - p) \geq 10$], tanto X como el estimador $\hat{p} = X/n$ están normalmente distribuidos en forma aproximada. Primero se consideran pruebas con muestras grandes basadas en este último hecho y luego se acude al caso de muestra pequeña que usa de modo directo la distribución binomial.

Pruebas con muestra grande

Las pruebas con muestra grande relacionadas con p son un caso especial de los procedimientos con muestra grande para un parámetro θ. Sea $\hat{\theta}$ un estimador de θ que es (por lo menos de manera aproximada) insesgado y que tiene aproximadamente una distribución normal. La hipótesis nula tiene la forma H_0: $\theta = \theta_0$, donde θ_0 denota un número (el valor nulo) apropiado al contexto del problema. Suponga que cuando H_0 es verdadera, la desviación estándar de $\hat{\theta}$, $\sigma_{\hat{\theta}}$, no implica parámetros desconocidos. Por ejemplo, si $\theta = \mu$ y $\hat{\theta} = \overline{X}$, $\sigma_{\hat{\theta}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$, la cual no implica parámetros desconocidos sólo si se conoce el valor de σ. Al estandarizar $\hat{\theta}$ conforme a la suposición de que H_0 es verdadera (de modo que $E(\hat{\theta}) = \theta_0$) se obtiene un estadístico de prueba para muestra grande:

$$\text{Estadístico de prueba: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

Si la hipótesis alternativa es H_a: $\theta > \theta_0$, la región de rechazo $z \geq z_\alpha$ especifica una prueba de cola superior cuyo nivel significativo es aproximadamente α. Las otras dos alternativas, H_a: $\theta < \theta_0$ y H_a: $\theta \neq \theta_0$, se someten a una prueba z de cola inferior y a una prueba z de dos colas, respectivamente.

En el caso $\theta = p$, $\sigma_{\hat{\theta}}$ no implicará parámetros desconocidos cuando H_0 es verdadera, aunque esto es atípico. Cuando $\sigma_{\hat{\theta}}$ implica parámetros desconocidos, a menudo es posible utilizar una desviación estándar estimada $S_{\hat{\theta}}$ en lugar de $\sigma_{\hat{\theta}}$ y seguir teniendo Z aproximadamente distribuida de manera normal cuando H_0 es verdadera (porque cuando n es grande $s_{\hat{\theta}} \approx \sigma_{\hat{\theta}}$ para la mayoría de las muestras). La prueba con muestra grande de la sección previa da un ejemplo de esto, como σ casi siempre es desconocida, se utiliza $s_{\hat{\theta}} = s_{\overline{X}} = s/\sqrt{n}$ en lugar de $\sigma/\sqrt{n}$ en el denominador de z.

El estimador $\hat{p} = X/n$ es $(E(\hat{p}) = p)$ insesgado y su distribución es aproximadamente normal y su desviación estándar es $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. Estos hechos se utilizaron en la sección 7.2 para obtener un intervalo de confianza para p. Cuando H_0 es verdadera, $E(\hat{p}) = p_0$ y $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n}$, $\sigma_{\hat{p}}$ no implica parámetros desconocidos. Se concluye entonces que cuando n es grande y H_0 es verdadera, el estadístico de prueba

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

tiene aproximadamente una distribución normal estándar. Si la hipótesis alternativa es H_a: $p > p_0$ y se utiliza la región de rechazo de cola superior $z \geq z_\alpha$, entonces

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(Z \geq z_\alpha \text{ cuando } Z \text{ tiene aproximadamente una distribución} \\ &\quad\ \text{normal estándar} \approx \alpha \end{aligned}$$

Por consiguiente, el nivel de significancia deseado α se obtiene utilizando el valor crítico que capture el área α en la cola superior de la curva z. Las regiones de rechazo para las otras dos hipótesis alternativas, cola inferior para H_a: $p < p_0$ y dos colas para H_a: $p \neq p_0$ se justifican de manera análoga.

Hipótesis nula: H_0: $p = p_0$

Valor del estadístico de prueba: $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

Hipótesis alternativa	Región de rechazo
H_a: $p > p_0$	$z \geq z_\alpha$ (cola superior)
H_a: $p < p_0$	$z \leq -z_\alpha$ (cola inferior)
H_a: $p \neq p_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ (dos colas)

Estos procedimientos de prueba son válidos siempre que $np_0 \geq 10$ y $n(1-p_0) \geq 10$.

Ejemplo 8.11 El corcho natural en botellas de vino está sujeto a deterioro y como resultado el vino de esas botellas puede experimentar contaminación. El artículo “Effects of Bottle Closure Type on Consumer Perceptions of Wine Quality” (*Amer. J. of Enology and Viticulture*, 2007: 182–191) informó que, en una degustación de chardonnay comerciales, 16 de 91 botellas se consideraron echadas a perder en cierta medida por las características asociadas del corcho. ¿Estos datos proporcionan una fuerte evidencia para concluir que más del 15% de todas las botellas están contaminadas de esta manera? Se realizará una prueba de hipótesis con un nivel de significancia de .10.

1. p = la verdadera proporción de todas las botellas de chardonnay comerciales consideradas inservibles en cierta medida por las características asociadas del corcho.
2. La hipótesis nula es H_0: $p = .15$.
3. La hipótesis alternativa es H_a: $p > .15$, la afirmación de que el porcentaje de población supera el 15%.

4. Como $np_0 = 91(.15) = 13.65 > 10$ y $nq_0 = 91(.85) = 77.35 > 10$, la prueba z con muestra grande puede ser utilizada. El valor estadístico de prueba es $z = (\hat{p} - .15)/\sqrt{(.15)(.85)/n}$.

5. La forma de H_a implica que una prueba de cola superior es apropiada: rechazar H_0 si $z \geq z_{.10} = 1.28$.

6. $\hat{p} = 16/91 = .1758$, de donde se obtiene

$$z = (.1758 - .15)/\sqrt{(.15)(.85)/91} = .0258/.0374 = .69$$

7. Como $.69 < 1.28$, z no está en la región de rechazo. Al nivel de significancia .10, la hipótesis nula no puede rechazarse. Aunque el porcentaje de botellas contaminadas en la muestra es un poco superior al 15%, el porcentaje de la muestra no es lo suficientemente grande como para concluir que el porcentaje de la población supera el 15%. La diferencia entre la proporción de la muestra .1758 y el valor nulo .15 puede ser explicado adecuadamente por la variabilidad de muestreo. ■

β y determinación del tamaño de la muestra Cuando H_0 es verdadera, el estadístico de prueba Z tiene aproximadamente una distribución normal estándar. Supóngase ahora que H_0 *no* es verdadera y que $p = p'$. Entonces Z sigue teniendo aproximadamente una distribución normal (porque es una función lineal de $\hat{p}$), pero su valor medio y su varianza ya no son 0 y 1, respectivamente. En su lugar,

$$E(Z) = \frac{p' - p_0}{\sqrt{p_0(1 - p_0)/n}} \qquad V(Z) = \frac{p'(1 - p')/n}{p_0(1 - p_0)/n}$$

La probabilidad de un error de tipo II para una prueba de cola superior es $\beta(p') = P(Z > z_\alpha$ cuando $p = p')$. Esto se puede calcular usando la media y la varianza dadas para estandarizar y entonces referirse a la función de distribución acumulativa normal estándar. Además, si se desea que el nivel de prueba α también tenga $\beta(p') = \beta$ para un valor especificado de β, esta ecuación se resuelve para la n necesaria como en la sección 8.2. En el recuadro adjunto se dan expresiones generales para $\beta(p')$ y n.

Hipótesis alternativa	$\beta(p')$
H_a: $p > p_0$	$\Phi\left[\dfrac{p_0 - p' + z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$
H_a: $p < p_0$	$1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$
H_a: $p \neq p_0$	$\Phi\left[\dfrac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right] - \Phi\left[\dfrac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$

El tamaño de muestra n con el cual la prueba de nivel α también satisface $\beta(p') = \beta$ es

$$n = \begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1 - p_0)} + z_\beta\sqrt{p'(1 - p')}}{p' - p_0}\right]^2 & \text{prueba de una cola} \\ \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1 - p_0)} + z_\beta\sqrt{p'(1 - p')}}{p' - p_0}\right]^2 & \text{prueba de dos colas (una solución aproximada)} \end{cases}$$

Ejemplo 8.12 Un servicio de mensajería anuncia que por lo menos 90% de todos los paquetes llevados a su oficina alrededor de las 9 a.m. para entrega en la misma ciudad son entregados alrededor del mediodía de ese mismo día. Sea p la proporción verdadera de dichos paquetes que son entregados como se anuncia y considere las hipótesis H_0: $p = .9$ contra H_a: $p < .9$. Si sólo 80% de los paquetes son entregados como se anuncia, ¿qué tan probable es que una prueba de nivel .01 basada en $n = 225$ paquetes detectará tal alejamiento de H_0? ¿Cuál debe ser el tamaño de la muestra para garantizar que $\beta(.8) = .01$? Con $\alpha = .01$, $p_0 = .9$, $p' = .8$ y $n = 225$,

$$\beta(.8) = 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}}\right)$$
$$= 1 - \Phi(2.00) = .0228$$

Así pues la probabilidad de que H_0 sea rechazada si se realiza la prueba cuando $p = .8$ es .9772, aproximadamente 98% de todas las muestras darán por resultado el rechazo correcto de H_0.

Con $z_\alpha = z_\beta = 2.33$ en la fórmula del tamaño de la muestra se obtiene

$$n = \left[\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9}\right]^2 \approx 266 \qquad \blacksquare$$

Pruebas con muestra pequeña

Los procedimientos de prueba cuando el tamaño de muestra n es pequeño están basados directamente en la distribución binomial en lugar de en la aproximación normal. Considérese la hipótesis alternativa H_a: $p > p_0$ y de nuevo sea X el número de éxitos en la muestra. Entonces X es el estadístico de prueba y la región de rechazo de cola superior tiene la forma $x \geq c$. Cuando H_0 es verdadera, X tiene una distribución binomial con parámetros n y p_0, por lo tanto

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(X \geq c \text{ cuando } X \sim \text{Bin}(n, p_0)) \\ &= 1 - P(X \leq c - 1 \text{ cuando } X \sim \text{Bin}(n, p_0) \\ &= 1 - B(c - 1; n, p_0) \end{aligned}$$

A medida que el valor crítico c disminuye, más valores x están incluidos en la región de rechazo y P(error de tipo I) se incrementa. Como X tiene una distribución de probabilidad discreta, por lo general no es posible hallar un valor de c con el cual P(error de tipo I) sea exactamente el nivel de significancia α deseado (p. ej., .05 o .01). En su lugar, se utiliza la región de rechazo más grande de la forma $\{c, c + 1, \ldots, n\}$ que satisface $1 - B(c - 1; n, p_0) \leq \alpha$.

Sea p' un valor alternativo de $p (p' > p_0)$. Cuando $p = p'$, $X \sim \text{Bin}(n, p')$, por lo tanto

$$\begin{aligned} \beta(p') &= P(\text{error de tipo II cuando } p = p') \\ &= P(X < c \text{ cuando } X \sim \text{Bin}(n, p')) = B(c - 1; n, p') \end{aligned}$$

Es decir, $\beta(p')$ es el resultado de un cálculo de probabilidad binomial directo. El tamaño de muestra n necesario para garantizar que una prueba de nivel α tiene una β especificada con valor alternativo particular p' debe ser determinado mediante ensayo y error utilizando la función de distribución acumulativa binomial.

Los procedimientos de prueba para H_a: $p < p_0$ y para H_a: $p \neq p_0$ se construyen de manera similar. En el primer caso, la región de rechazo apropiada tiene la forma $x \leq c$ (una prueba de cola inferior). El valor crítico c es el número más grande que satisface $B(c; n, p_0) \leq \alpha$. La región de rechazo cuando la hipótesis alternativa es H_a: $p \neq p_0$ se compone tanto de valores x grandes como pequeños.

Ejemplo 8.13 Un fabricante de plástico desarrolló un nuevo tipo de bote de plástico para la basura y propone venderlo con una garantía incondicional de 6 años. Para ver si esto es factible desde el punto de vista económico, 20 botes prototipo se someten a una prueba acelerada de duración para simular 6 años de uso. La garantía propuesta se modificará sólo si los datos muestrales sugieren fuertemente que menos del 90% de los botes sobrevivirían el periodo de 6 años. Sea p la proporción de todos los botes que sobreviven la prueba acelerada. Las hipótesis pertinentes son H_0; $p = .9$ contra H_a: $p < .9$. La decisión se basará en el estadístico de prueba X, el número de entre los 20 que sobreviven. Si el nivel de significancia deseado es $\alpha = .05$, c debe satisfacer $B(c; 20, .9) \leq .05$. De acuerdo con la tabla A.1 del apéndice, $B(15; 20, .9) = .043$, mientras que $B(16; 20, .9) = .133$. La región de rechazo apropiada es por consiguiente $x \leq 15$. Si la prueba acelerada da por resultado $x = 14$, H_0 sería rechazada a favor de H_a y se modificaría la garantía propuesta. La probabilidad de un error de tipo II con el valor alternativo $p' = .8$ es

$$\begin{aligned}\beta(.8) &= P(H_0 \text{ no es rechazada cuando } X \sim \text{Bin}(20, .8)) \\ &= P(X \geq 16 \text{ cuando } X \sim \text{Bin}(20, .8)) \\ &= 1 - B(15; 20, .8) = 1 - .370 = .630\end{aligned}$$

Es decir, cuando $p = .8$, 63% de todas las muestras compuestas de $n = 20$ botes darían por resultado que H_0 sea incorrectamente no rechazada. Esta probabilidad de error es alta porque 20 es un tamaño de muestra pequeño y $p' = .8$ se acerca al valor nulo $p_0 = .9$. ■

EJERCICIOS Sección 8.3 (37–46)

37. Una caracterización común de las personas obesas es que su índice de masa corporal es de al menos 30 [IMC = peso/(altura)^2, donde la altura está en metros y el peso en kilogramos]. El artículo "The Impact of Obesity on Illness Absence and Productivity in an Industrial Population of Petrochemical Workers" (*Annals of Epidemiology*, 2008: 8−14) informó que en una muestra de mujeres trabajadoras, 262 tenían un índice de masa corporal inferior a 25, 159 tenían IMC que era al menos 25 pero no más de 30 y 120 tenían un IMC superior a 30. ¿Hay pruebas contundentes para concluir que más del 20% de individuos en la población analizada son obesos?

a. Establezca las hipótesis de prueba adecuadas utilizando el enfoque de región de rechazo con un nivel de significancia de .05.

b. Explique en el contexto de este escenario qué constituye errores de tipos I y II.

c. ¿Cuál es la probabilidad de no concluir que más del 20% de la población es obesa cuando el porcentaje real de los individuos obesos es del 25%?

38. Un fabricante de baterías de níquel-hidrógeno selecciona al azar 100 placas de níquel para probar las celdas, someterlas a ciclos un número especificado de veces y concluye que 14 de ellas se ampollan en tales circunstancias.

a. ¿Proporciona esto una evidencia precisa para concluir que más de 10% de todas las placas se ampollan en tales circunstancias? Formule y pruebe las hipótesis apropiadas con un nivel de significancia de .05. Al llegar a su conclusión, ¿qué tipo de error pudo haber cometido?

b. Si es realmente el caso que 15% de todas las placas se ampollan en estas circunstancias y se utiliza un tamaño de muestra de 100, ¿qué tan probable es que la hipótesis nula del inciso (a) no sea rechazada por la prueba de nivel .05? Responda esta pregunta para un tamaño de muestra de 200.

c. ¿Cuántas placas tendrían que ser probadas para tener $\beta(.15) = .10$ para la prueba del inciso (a)?

39. Una muestra aleatoria de 150 donaciones recientes en un banco de sangre revela que 82 fueron de sangre tipo A. ¿Sugiere esto que el porcentaje real de donaciones tipo A difiere de 40%, el porcentaje de la población que tiene sangre tipo A? Realice una prueba de las hipótesis apropiadas utilizando un nivel de significancia de .01. ¿Habría sido diferente su conclusión si se hubiera utilizado un nivel de significancia de .05?

40. Se sabe que aproximadamente 2/3 de todos los seres humanos tienen un ojo o pie derecho dominantes. ¿Existe también dominio del lado derecho en el acto de besar? El artículo "Human Behavior: Adult Persistence of Head-Turning Asymmetry" (*Nature*, 2003: 771) reportó que en una muestra aleatoria de 124 parejas que se besan, ambas personas en 80 de las parejas tendieron a inclinarse más hacia la derecha que hacia la izquierda.

a. Si 2/3 de las parejas que se besan exhiben esta tendencia de inclinarse hacia la derecha, ¿cuál es la probabilidad de que el número en una muestra de 124 que lo hacen así difiera del valor esperado en por lo menos lo que en realidad se observó?

b. ¿Sugiere este resultado del experimento que la cifra de 2/3 es no factible como comportamiento al besarse? Formule y pruebe las hipótesis apropiadas.

41. El artículo referido en el ejemplo 8.11 también informó que en una muestra de 106 consumidores de vino, 22 (20.8%) opina que las tapas de tornillo son un sustituto aceptable para los corchos naturales. Suponga que una bodega particular decide utilizar tapas de rosca en uno de sus vinos a menos que haya

Respuestas a ejercicios seleccionados de número impar

Capítulo 1

1. **a.** *Los Angeles Times, Oberlin Tribune, Gainesville Sun, Washington Post*
   b. Duke Energy, Clorox, Seagate, Neiman Marcus
   c. Vince Correa, Catherine Miller, Michael Cutler, Ken Lee
   d. 2.97, 3.56, 2.20, 2.97

3. **a.** ¿Qué tan probable es que más de la mitad de las computadoras muestreadas necesiten o hayan necesitado servicio de garantía? ¿Cuál es el número esperado entre los 100 que necesitan servicio de garantía? ¿Qué tan probable es que el número que necesita servicio de garantía sea mayor que el número esperado en más de 10?
   b. Suponga que 15 de las 100 muestreadas necesitaba servicio de garantía. ¿Qué confianza podemos tener de que la proporción de *todas* las computadoras que necesitan servicio de garantía sea entre .08 y .22? ¿La muestra proporciona evidencia obligatoria para concluir que más del 10% de esas computadoras necesitan servicio de garantía?

5. **a.** No. Todos los estudiantes que toman un curso grande de estadística que participan en un programa de IS de esta clase.
   b. La aleatoriedad protege contra sesgos diversos, y ayuda a asegurar que los del grupo IS son tan semejantes como es posible a los estudiantes del grupo de control.
   c. No habría base firme para evaluar la efectividad de la IS (nada a que las calificaciones de la IS pudieran compararse razonablemente).

7. Uno podría generar una muestra aleatoria sencilla de todas las casas unifamiliares de la ciudad, o una muestra aleatoria estratificada al tomar una muestra aleatoria sencilla de cada uno de los 10 vecindarios del distrito. De cada una de las casas seleccionadas, se determinarían valores de todas las variables deseadas. Éste sería un estudio enumerativo porque existe una población de objetos finita e identificable de la cual tomar muestras.

9. **a.** Posiblemente por error de medición, error de registro, diferencias en condiciones ambientales en el momento de la medición, etc.
   b. No. No hay marco de muestreo.

11.

6L	430
6H	769689
7L	42014202
7H	
8L	011211410342
8H	9595578
9L	30
9H	58

La brecha en los datos, no hay calificaciones en los 70 altos.

13. **a.** 12 2 hoja: dígito de unidades
12 445
12 6667777
12 889999
13 00011111111
13 2222222222333333333333333
13 44444444444444444455555555555555555555
13 6666666666667777777777
13 888888888888999999
14 0000001111
14 2333333
14 444
14 77
simetría
   b. Cerca de la forma de campana, centro ≈ 135, no hay dispersión insignificante, no hay brechas ni valores atípicos.

15.

Am		Fr
	8	1
157020153504	9	00645632
9324	10	2563
6306	11	6913
058	12	325528
8	13	7
	14	
	15	8
2	16	

Tallo: centenas y decenas
Hoja: unidades

Valores representativos: debajo de 100 para Am y debajo de 110 para Fr. Algo más de variabilidad en los tiempos de Fr que en horarios de Am. Sesgo positivo más extremo de Am que para Fr. 162 es un valor atípico Am y 158 es quizás un valor atípico para Fr.

17. a.

# No se apega	Frecuencia	Frec. rel.
0	7	.117
1	12	.200
2	13	.217
3	14	.233
4	6	.100
5	3	.050
6	3	.050
7	1	.017
8	1	.017
	60	1.001

b. .917, .867, 1 − .867 = .133

c. El histograma tiene un sesgo positivo considerable. Está centrado en un punto entre 2 y 3 y se dispersa bastante alrededor de su centro.

19. a. .99 (99%), .71 (71%) **b.** .64 (64%), .44 (44%)

c. Estrictamente hablando, el histograma no es unimodal, pero está cerca de serlo con un sesgo positivo moderado. Es probable que un tamaño muestral mucho más grande diera una imagen más uniforme.

21. a.

y	Frec.	Frec. rel.
0	17	.362
1	22	.468
2	6	.128
3	1	.021
4	0	.000
5	1	.021
	47	1.000

.362, .638

b.

z	Frec.	Frec. rel.
0	13	.277
1	11	.234
2	3	.064
3	7	.149
4	5	.106
5	3	.064
6	3	.064
7	0	.000
8	2	.043
	47	1.001

.894, .830

23. Los anchos de clase no son iguales, por lo que la escala de densidad debe ser utilizada. Las densidades de las seis clases son .2030, .1373, .0303, .0086, .0021 y .0009, respectivamente. El histograma resultante es unimodal con un sesgo positivo muy sustancial.

25.

Clase	Frec.	Clase	Frec.
10–<20	8	1.1–<1.2	2
20–<30	14	1.2–<1.3	6
30–<40	8	1.3–<1.4	7
40–<50	4	1.4–<1.5	9
50–<60	3	1.5–<1.6	6
60–<70	2	1.6–<1.7	4
70–<80	1	1.7–<1.8	5
	40	1.8–<1.9	1
			40

Original: sesgado positivamente;
Transformado: mucho más simétrico, no lejos de forma de campana.

27. a. La observación 50 cae en un límite de clase.

b.

Clase	Frec.	Frec. rel.
0–<50	9	.18
50–<100	19	.38
100–<150	11	.22
150–<200	4	.08
200–<300	4	.08
300–<400	2	.04
400–<500	0	.00
500–<600	1	.02
	50	1.00

Un valor representativo (central) está un poco abajo o un poco arriba de 100, dependiendo de cómo se mida el centro. Hay gran variabilidad en los tiempos de vida, en especial en valores del extremo superior de los datos. Hay varios candidatos para valores apartados.

c.

Clase	Frec.	Frec. rel.
2.25–<2.75	2	.04
2.75–<3.25	2	.04
3.25–<3.75	3	.06
3.75–<4.25	8	.16
4.25–<4.75	18	.36
4.75–<5.25	10	.20
5.25–<5.75	4	.08
5.75–<6.25	3	.06
	50	1.00

Hay mucha más simetría en la distribución de los valores $\ln(x)$ que en los valores x en sí, y menos variabilidad. Ya no hay brechas ni valores apartados obvios.

d. .38, .14

29.

Queja	Frec.	Frec. rel.
J	10	.1667
F	9	.1500
B	7	.1167
M	4	.0667
C	3	.0500
N	6	.1000
O	21	.3500
	60	1.0001

31.

Clase	Frec.	Frec. acum.	Frec. acumul. rel.
0–<4	2	2	.050
4–<8	14	16	.400
8–<12	11	27	.675
12–<16	8	35	.875
16–<20	4	39	.975
20–<24	0	39	.975
24–<28	1	40	1.000

33. a. 640.5, 582.5
b. 610.5, 582.5
c. 591.2
d. 593.71

35. a. $\bar{x} = 12.55$, $\tilde{x} = 12.50$, $\bar{x}_{tr(12.5)} = 12.40$. Borrar la observación más grande (18.0) hace que $\tilde{x}$ y $\bar{x}_{tr}$ sean un poco menores que $\bar{x}$.
b. En casi 4.0 **c.** No; multiplique los valores de $\bar{x}$ y $\tilde{x}$ por el factor de conversión 1/2.2.

37. $\bar{x}_{tr(10)} = 11.46$

39. a. $\bar{x} = 1.0297$, $\tilde{x} = 1.009$ **b.** .383

41. a. .7 **b.** También .7 **c.** 13

43. $\tilde{x} = 68.0$, $\bar{x}_{tr(20)} = 66.2$, $\bar{x}_{tr(30)} = 67.5$

45. a. $\bar{x} = 115.58$; las desviaciones son .82, .32, $-.98$, $-.38$, .22
b. .482, .694 **c.** .482 **d.** .482

47. $\bar{x} = 116.2$, $s = 25.75$. La magnitud de s indica una cantidad importante de variación alrededor del centro (una desviación "representativa" de casi 25).

49. a. 56.80, 197.8040 **b.** .5016, .708

51. a. 1264.766, 35.564 **b.** .351, .593

53. a. Bal: 1.121, 1.050, .536
Gr: 1.244, 1.100, .448
b. Proporciones típicas son muy similares para los dos tipos. Hay una variabilidad algo más en la muestra Bal, debido principalmente a los dos valores extremos (uno leve, un extremo). Para Bal, hay simetría sustancial en el centro 50%, pero la asimetría en general positivo. Para los de Gr, hay sesgo positivo sustancial en el centro 50% y un leve sesgo positivo en general.

55. a. 33 **b.** No
c. Un ligero sesgo positivo en la mitad central, pero más bien simétrico en su conjunto. La magnitud de variabilidad parece importante.
d. A lo sumo 32

57. a. Sí. 125.8 es un valor atípico extremo y 250.2 es un valor atípico moderado.
b. Además de la presencia de valores atípicos, hay sesgo positivo tanto en el 50% central de los datos como en general, excepto los valores atípicos. Con excepción de los dos valores atípicos, parece haber una cantidad relativamente pequeña de variabilidad en los datos.

59. a. DE (delirio excitado): .4, .10, 2.75, 2.65;
No-DE: 1.60, .30, 7.90, 7.60
b. DE: 8.9 y 9.2 son valores apartados moderados, y 11.7 y 21.0 son valores atípicos extremos.
No hay valores atípicos en la muestra sin delirio excitado.
c. Cuatro valores atípicos para DE, ninguno para no-DE. Sesgo positivo importante en ambas muestras; menos variabilidad en DE (para f_s menor), y las observaciones para no-DE tienden a ser un poco mayores que las observaciones con DE.

61. Valores atípicos, moderados y extremos, sólo a las 6 a.m. Las distribuciones a otras horas son bastante simétricas. La variabilidad aumenta un poco hasta las 2 p.m. y luego disminuye ligeramente, y lo mismo es cierto de valores "típicos" de coeficiente de vapor de gasolina.

63. $\bar{x} = 64.89$, $\tilde{x} = 64.70$, $s = 7.803$, baja 4ª = 57.8, alta 4ª = 70.4, $f_s = 12.6$. Un histograma que consta de 8 clases a partir de las 52, cada una de ancho de 4, es bimodal, pero cerca de unimodal con un sesgo positivo. Un diagrama de caja no muestra los valores extremos, hay un sesgo negativo muy leve en el centro 50% y el bigote superior es mucho más largo que el bigote más bajo.
b. .9231, .9053
c. .48

67. a. M: $\bar{x} = 3.64$, $\tilde{x} = 3.70$, $s = .269$, $f_s = .40$
F: $\bar{x} = 3.28$, $\tilde{x} = 3.15$, $s = .478$, $f_s = .50$
Los valores femeninos son por lo general un poco menores que los masculinos, y muestran un poco de más variabilidad. Una gráfica de caja M muestra sesgo negativo, mientras que una gráfica de caja F muestra sesgo positivo.
b. F: $\bar{x}_{tr(10)} = 3.24$ M: $\bar{x}_{tr(10)} = 3.652 \approx 3.65$

69. a. $\bar{y} = a\bar{x} + b$, $s_y^2 = a^2 s_x^2$ **b.** 189.14, 1.87

71. a. La media, mediana y media recortada son prácticamente idénticas, lo que sugiere una cantidad importante de simetría en la información; el hecho que los cuartiles estén a casi la misma distancia desde la mediana, y que las observaciones máximas sean casi equidistantes del centro, proporciona más apoyo para simetría. La desviación estándar es bastante pequeña con respecto a la media y mediana.
b. Vea los comentarios de (a). Además, usando 1.5(Q3 − Q1) como medida, las dos observaciones más grandes y las tres más pequeñas son valores atípicos moderados.

73. $\bar{x} = .9255$, $s = .0809$, $\tilde{x} = .93$, pequeña cantidad de variabilidad, un poco de sesgo

75. a. Los "resúmenes de cinco números" ($\tilde{x}$, los dos cuartos y las observaciones más pequeña y más grande) son idénticas y no hay valores apartados, de modo que las tres gráficas de caja individuales son idénticas.
b. Diferencias en variabilidad, naturaleza de brechas, y existencia de grupos para tres muestras.
c. No. Se pierde el detalle.

77. c. Las profundidades representativas son muy similares para los cuatro tipos de suelos, entre 1.5 y 2. Los datos de la C y suelos CL muestran mucho más variabilidad que los otros dos tipos. Los diagramas de caja para los tres primeros tipos muestran un importante sesgo positivo tanto en el centro 50% y en general. El diagrama de caja para el suelo SYCL muestra asimetría negativa en el centro 50% y un leve sesgo positivo en general. Por último, hay múltiples valores atípicos para los tres primeros tipos de suelos, incluyendo los valores atípicos extremos.

79. **a.** $\bar{x}_{n+1} = (n\bar{x}_n + x_{n+1})/(n + 1)$
c. 12.53, .532

81. Una asimetricidad positiva importante (suponiendo unimodalidad)

83. **a.** Todos los puntos caen en una recta de 45°. Los puntos caen debajo de la recta de 45°
b. Los puntos que caen debajo de la recta de 45°, indican una asimetricidad positiva importante.

Capítulo 2

1. **a.** $\mathcal{S}$ = {1324, 3124, 1342, 3142, 1423, 1432, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
b. A = {1324, 1342, 1423, 1432}
c. B = {2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
d. $A \cup B$ = {1324, 1342, 1423, 1432, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231},
$A \cap B$ no contiene resultados (A y B son disjuntos),
A' = {3124, 3142, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}

3. **a.** A = {SSF, SFS, FSS}
b. B = {SSF, SFS, FSS, SSS}
c. C = {SFS, SSF, SSS}
d. C' = {FFF, FSF, FFS, FSS, SFF},
$A \cup C$ = {SSF, SFS, FSS, SSS},
$A \cap C$ = {SSF, SFS},
$B \cup C$ = {SSF, SFS, FSS, SSS} = B,
$B \cap C$ = {SSF, SFS, SSS} = C,

5. **a.** $\mathcal{S}$ = {(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 1), (1, 2, 2), (1, 2, 3), (1, 3, 1), (1, 3, 2), (1, 3, 3), (2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 2, 1), (2, 2, 2), (2, 2, 3), (2, 3, 1), (2, 3, 2), (2, 3, 3), (3, 1, 1), (3, 1, 2), (3, 1, 3), (3, 2, 1), (3, 2, 2), (3, 2, 3), (3, 3, 1), (3, 3, 2), (3, 3, 3)} **b.** {(1, 1, 1), (2, 2, 2), (3, 3, 3)} **c.** {(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)} **d.** {(1, 1, 1), (1, 1, 3), (1, 3, 1), (1, 3, 3), (3, 1, 1), (3, 1, 3), (3, 3, 1), (3, 3, 3)}

7. **a.** Hay 35 resultados en $\mathcal{S}$. **b.** {*AABABAB*, *AABAABB*, *AAABBAB*, *AAABABB*, *AAAABBB*}

11. **a.** .07 **b.** .30 **c.** .57

13. **a.** .36 **b.** .64 **c.** .53
d. .47 **e.** .17 **f.** .75

15. **a.** .572 **b.** .879

17. **a.** Hay paquetes computarizados de estadística que no son SPSS y SAS.
b. .70 **c.** .80 **d.** .20

19. **a.** .8841 **b.** .0435

21. **a.** .10 **b.** .18, .19 **c.** .41 **d.** .59
e. .31 **f.** .69

23. **a.** .067 **b.** .400 **c.** .933 **d.** .533

25. **a.** .85 **b.** .15 **c.** .22 **d.** .35

27. **a.** .1 **b.** .7 **c.** .6

29. **a.** 676; 1296 **b.** 17,576; 46,656
c. 456,976; 1,679,616 **d.** .942

31. **a.** 243 **b.** 3645 días (casi 10 años)

33. **a.** 1,816,214,400 **b.** 659,067,881,572,000 **c.** 9,072,000

35. **a.** 38,760, .0048 **b.** .0054 **c.** .9946 **d.** .2885

37. **a.** 60 **b.** 10 **c.** .0456

39. **a.** .0839 **b.** .24975

41. **a.** 10,000 **b.** .9876 **c.** .0333 **d.** .0337

43. .000394, .00394, .00001539

45. **a.** .447, .500, .200 **b.** .400, .447 **c.** .211

47. **a.** .50 **b.** .50 **c.** .6125
d. .3875 **e.** .769

49. **a.** .34, .40 **b.** .588 **c.** .50

51. **a.** .436, **b.** .581

53. .083

55. .236

59. **a.** .21 **b.** .455 **c.** .264, .462, .274

61. **a.** .578, .278, .144 **b.** 0, .457, .543

63. **b.** .54 **c.** .68 **d.** .74 **e.** .7941

65. .087, .652, .261

67. .000329; muy inquietante.

69. **a.** .126 **b.** .05 **c.** .1125 **d.** .2725
e. .5325 **f.** .2113

71. **a.** .300 **b.** .820 **c.** .146

75. .401, .723

77. **a.** .00889 **b.** .00421

79. .0059

81. **a.** .95

83. **a.** .10, .20 **b.** 0

85. **a.** $p(2 - p)$ **b.** $1 - (1 - p)^n$ **c.** $(1 - p)^3$
d. $.9 + (1 - p)^3(.1)$
e. $.1(1 - p)^3/[.9 + .1(1 - p)^3] = .0137$ para $p = .5$

87. **a.** .40 **b.** .571
c. No: $.571 \neq .65$, y también $.40 \neq (.65)(.7)$ **d.** .733

89. $[2\pi(1 - \pi)]/(1 - \pi^2)$

91. **a.** .333, .444 **b.** .150 **c.** .291

93. .45, .32

95. **a.** .0083 **b.** .2 **c.** .2 **d.** .1074

97. .905

99. **a.** .974 **b.** .9754

101. .926

103. **a.** .018 **b.** .601

105. **a.** .883, .117 **b.** 23 **c.** .156

107. $1 - (1 - p_1)(1 - p_2) \cdot \cdots \cdot (1 - p_n)$

109. **a.** .0417 **b.** .375

111. P(contratar #1) $= 6/24$ para $s = 0$, $= 11/24$ para $s = 1$, $= 10/24$ para $s = 2$, y $= 6/24$ para $s = 3$, de modo que $s = 1$ es mejor.

113. $1/4 = P(A_1 \cap A_2 \cap A_3)$
$\neq P(A_1) \cdot P(A_2) \cdot P(A_3) = 1/8$

Capítulo 3

1. $x = 0$ para *FFF;* $x = 1$ para *SFF, FSF,* y *FFS;* $x = 2$ para *SSF, SFS,* y *FSS;* y $x = 3$ para *SSS*

3. Z = promedio de los dos números, con posibles valores 2/2, 3/2, . . . , 12/2; W = valor absoluto de la diferencia, con valores posibles 0, 1, 2, 3, 4, 5

5. No. En el ejemplo 3.4, sea $Y = 1$ si a lo sumo se examinan 3 baterías y sea $Y = 0$ de otro modo. Entonces Y tiene sólo dos valores.

7. **a.** $\{0, 1, \ldots, 12\}$; discreta **c.** $\{1, 2, 3, \ldots\}$; discreta **e.** $\{0, c, 2c, \ldots, 10{,}000c\}$, donde c es la regalía por libro; discreta **g.** $\{x: m \leq x \leq M\}$ donde $m(M)$ es la tensión mínima (máxima) posible; continua

9. **a.** $\{2, 4, 6, 8, \ldots\}$, es decir, $\{2(1), 2(2), 2(3), 2(4), \ldots\}$, una sucesión infinita; discreta
b. $\{2, 3, 4, 5, 6, \ldots\}$, es decir, $\{1 + 1, 1 + 2, 1 + 3, 1 + 4, \ldots\}$, una sucesión infinita, discreta

11. **a.** $p(4) = .45$, $p(6) = .40$, $p(8) = .15$, $p(x) = 0$ para $x \neq 4, 6$, u 8 **c.** .55, .15

13. **a.** .70 **b.** .45 **c.** .55
d. .71 **e.** .65 **f.** .45

15. **a.** (1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5) **b.** $p(0) = .3$, $p(1) = .6$, $p(2) = .1$
c. $F(x) = 0$ para $x < 0$, $= .3$ para $0 \leq x < 1$, $= .9$ para $1 \leq x < 2$, e $= 1$ para $2 \leq x$

17. **a.** .81 **b.** .162 **c.** Es A; *AUUUA, UAUUA, UUAUA, UUUAA;* .0036

19. $p(0) = .09$, $p(1) = .40$, $p(2) = .32$, $p(3) = .19$

21. **b.** $p(x) = .301, .176, .125, .097, .079, .067, .058, .051, .046$ para $x = 1, 2, \ldots, 9$
c. $F(x) = 0$ para $x < 1$, $= .301$ para $1 \leq x < 2$, $= .477$ para $2 \leq x < 3, \ldots, = .954$ para $8 \leq x < 9$, $= 1$ para $x \geq 9$
d. .602, .301

23. **a.** .20 **b.** .33 **c.** .78 **d.** .53

25. **a.** $p(y) = (1 - p)^y \cdot p$ para $y = 0, 1, 2, 3, \ldots$

27. **a.** 1234, 1243, 1324, . . . , 4321
b. $p(0) = 9/24$, $p(1) = 8/24$, $p(2) = 6/24$, $p(3) = 0$, $p(4) = 1/24$

29. **a.** 6.45 **b.** 15.6475 **c.** 3.96 **d.** 15.6475

31. 4.49, 2.12, .68

33. **a.** p **b.** $p(1 - p)$ **c.** p

35. $E[h_3(X)] = \$4.93$, $E[h_4(X)] = \$5.33$, así que 4 copias es mejor.

37. $E(X) = (n + 1)/2$, $E(X^2) = (n + 1)(2n + 1)/6$, $V(X) = (n^2 - 1)/12$

39. 6.1, .81, 88.5, 20.25

43. $E(X - c) = E(X) - c$, $E(X - \mu) = 0$

47. **a.** .515 **b.** .219 **c.** .012 **d.** .480
e. .965 **f.** .000 **g.** .595

49. **a.** .354 **b.** .114 **c.** .918

51. **a.** 6.25 **b.** 2.17 **c.** .030

53. **a.** .403 **b.** .787 **c.** .774

55. .1478

57. .407, independencia

59. **a.** .017 **b.** .811, .425 **c.** .006, .902, .586

61. Cuando $p = .9$, la probabilidad es .99 para A y .9963 para B. Si $p = .5$, estas probabilidades son .75 y .6875, respectivamente.

63. La tabulación para $p > .5$ es innecesaria.

65. **a.** 20, 16 **b.** 70, 21

67. $P(|X - \mu| \geq 2\sigma) = .042$ cuando $p = .5$ e $= .065$ cuando $p = .75$, en comparación con el límite superior de .25. Usando $k = 3$ en lugar de $k = 2$, estas probabilidades son .002 y .004, respectivamente, mientras que el límite superior es .11.

69. **a.** .114 **b.** .879 **c.** .121 **d.** Use la distribución binomial con $n = 15$, $p = .10$

71. **a.** $h(x;\ 15, 10, 20)$ para $x = 5, \ldots, 10$ **b.** .0325 **c.** .697

73. **a.** $h(x;\ 10, 10, 20)$ **b.** .033 **c.** $h(x;\ n, n, 2n)$

75. **a.** $nb(x;\ 2, .5)$ **b.** .188 **c.** .688 **d.** 2, 4

77. $nb(x;\ 6, .5)$, 6

79. **a.** .932 **b.** .065 **c.** .068 **d.** .492 **e.** .251

81. **a.** .011 **b.** .441 **c.** .554, .459 **d.** .945

83. Poisson(5) **a.** .492 **b.** .133

85. **a.** .122, .809, .283 **b.** 12, 3.464 **c.** .530, .011

87. **a.** .099 **b.** .135 **c.** 2

89. **a.** 4 **b.** .215 **c.** Al menos $-\ln(.1)/2 \approx 1.1513$ años

91. **a.** .221 **b.** 6,800,000 **c.** $p(x;\ 20.106)$

95. **b.** 3.114, .405, .636

97. **a.** $b(x;\ 15, .75)$ **b.** .686 **c.** .313 **d.** 11.25, 2.81 **e.** .310

99. .991

101. **a.** $p(x;\ 2.5)$ **b.** .067 **c.** .109

103. 1.813, 3.05

105. $p(2) = p^2$, $p(3) = (1 - p)p^2$, $p(4) = (1 - p)p^2$, $p(x) = [1 - p(2) - \cdots - p(x - 3)](1 - p)p^2$ para $x = 5, 6, 7, \ldots$; .99950841

107. **a.** .0029 **b.** .0767, .9702

109. **a.** .135 **b.** .00144 **c.** $\sum_{x=0}^{\infty}[p(x;\ 2)]^5$

111. 3.590

113. **a.** No **b.** .0273

115. **b.** $.5\mu_1 + .5\mu_2$ **c.** $.25(\mu_1 - \mu_2)^2 + .5(\mu_1 + \mu_2)$ **d.** .6 y .4 sustituya .5 y .5, respectivamente.

117. $\sum_{i=1}^{10}(p_{i+j+1} + p_{i-j-1})p_i$, donde $p_k = 0$ si $k < 0$ o $k > 10$.

121. **a.** 2.50 **b.** 3.1

Capítulo 4

1. **b.** .4625 **c.** .5, .2781

3. **b.** .5 **c.** .6875 **d.** .6328

5. **a.** .375 **b.** .125 **c.** .297 **d.** .578

7. **a.** $f(x) = .1$ para $25 \le x \le 35$ y 0 de otro modo **b.** .20 **c.** .40 **d.** .20

9. **a.** .562 **b.** .438, .438 **c.** .071

11. **a.** .25 **b.** .1875 **c.** .4375 **d.** 1.4142 **e.** $f(x) = x/2$ para $0 < x < 2$ **f.** 1.33 **g.** .222, .471 **h.** 2

13. **a.** 3 **b.** 0 para $x \le 1$, $1 - x^{-3}$ para $x > 1$ **c.** .125, .088 **d.** 1.5, .866 **e.** .924

15. **a.** $F(x) = 0$ para $x \le 0$, $= 90\left[\frac{x^9}{9} - \frac{x^{10}}{10}\right]$ para $0 < x < 1$, $= 1$ para $x \ge 1$ **b.** .0107 **c.** .0107, .0107 **d.** .9036 **e.** .818, .111 **f.** .3137

17. **a.** $A + (B - A)p$ **b.** $E(X) = (A + B)/2$, $\sigma_X = (B - A)/\sqrt{12}$ **c.** $[B^{n+1} - A^{n+1}]/[(n + 1)(B - A)]$

19. **a.** .597 **b.** .369 **c.** $f(x) = .3466 - .25\ln(x)$ para $0 < x < 4$

21. 314.79

23. 248, 3.60

25. **b.** 1.8(90avo percentil para X) + 32 **c.** a(Xpercentil) + b

27. 0, 1.814

29. **a.** 2.14 **b.** .81 **c.** 1.17 **d.** .97 **e.** 2.41

31. **a.** 2.54 **b.** 1.34 **c.** $-.42$

33. **a.** .9664 **b.** .2451 **c.** .8664

35. **a.** .3336 **b.** Aproximadamente 0 **c.** .5795 **d.** 6.524 **e.** .8028

37. **a.** 0, .5793, .5793 **b.** .3174, no **c.** < 87.6 o > 120.4

39. **a.** 36.7 **b.** 22.225 **c.** 3.179

41. .002

43. 10, .2

45. 7.3%

47. 21.155

49. **a.** .1190, .6969 **b.** .0021 **c.** .7054 **d.** > 5020 o < 1844 (usando $z_{.0005} = 3.295$) **e.** Normal, $\mu = 7.576$, $\sigma = 1.064$, .7054

51. .3174 para $k = 1$, .0456 para $k = 2$, .0026 para $k = 3$, en comparación con los límites de 1, .25, y .111, respectivamente.

53. **a.** Exacta: .212, .577, .573; Aproximada: .211, .567, .596
b. Exacta: .885, .575, .017; Aproximada: .885, .579, .012
c. Exacta: .002, .029, .617; Aproximada: .003, .033, .599

55. **a.** .9409 **b.** .9943

57. **b.** Normal, $\mu = 239$, $\sigma^2 = 12.96$

59. **a.** 1 **b.** 1 **c.** .982 **d.** .129

61. **a.** .480, .667, .147 **b.** .050, 0

63. **a.** corto $\Rightarrow$ plan #1 mejor, mientras que largo $\Rightarrow$ plan #2mejor
b. $1/\lambda = 10 \Rightarrow E[h_1(X)] = 100$, $E[h_2(X)] = 112.53$
$1/\lambda = 15 \Rightarrow E[h_1(X)] = 150$, $E[h_2(X)] = 138.51$

65. **a.** .238 **b.** .238 **c.** .313 **d.** .653 **e.** .653
f. .713

67. **a.** .424 **b.** .567, $\tilde{\mu} < 24$ **c.** 60 **d.** 66

69. **a.** $\cap A_i$ **b.** Exponencial con $\lambda = .05$
c. Exponencial con parámetro $n\lambda$

73. **a.** .826, .826, .0636 **b.** .664 **c.** 172.727

77. **a.** 123.97, 117.373 **b.** .5517 **c.** .1587

79. **a.** 9.164, .385 **b.** .8790 **c.** .4247, asimetricidad
d. No, ya que $P(X < 17{,}000) = .9332$

81. **a.** 149.157, 223.595 **b.** .9573 **c.** .0414
d. 148.41 **e.** 9.57 **f.** 125.90

83. $\alpha = \beta$

85. **b.** $[\Gamma(\alpha + \beta) \cdot \Gamma(m + \beta)]/[\Gamma(\alpha + \beta + m) \cdot \Gamma(\beta)]$,
$\beta/(\alpha + \beta)$

87. Sí, porque el patrón en la gráfica es bastante lineal.

89. Sí

91. Sí

93. Graficar $\ln(x)$ vs. percentil z. El patrón es recto, de modo que es posible una distribución poblacional lognormal.

95. El patrón en la gráfica es bastante lineal; es muy posible que la resistencia sea distribuida normalmente.

97. Hay una curvatura importante en la gráfica. λ es un parámetro de escala (como es σ para la familia normal).

99. **a.** $F(y) = \frac{1}{48}(y^2 - y^3/18)$ para $0 \le y \le 12$
b. .259, .5, .241 **c.** 6, 43.2, 7.2
d. .518 **e.** 3.75

101. **a.** $f(x) = x^2$ para $0 \le x < 1$ e $= \frac{7}{4} - \frac{3}{4}x$ para $1 \le x \le \frac{7}{3}$ **b.** .917 **c.** 1.213

103. **a.** .9162 **b.** .9549 **c.** 1.3374

105. **a.** .3859 **b.** .1266 **c.** (72.97, 119.03)

107. **b.** $F(x) = 0$ para $x < -1$, $= (4x - x^3/3)/9 + \frac{11}{27}$ para $-1 \le x \le 2$, e $= 1$ para $x > 2$
c. No. $F(0) < .5 \Rightarrow \tilde{\mu} > 0$
d. $Y \sim \text{Bin}(10, \frac{5}{27})$

109. **a.** .368, .828, .460 **b.** 352.53
c. $1/\beta \cdot \exp[-\exp(-(x - \alpha)/\beta)] \cdot \exp(-(x - \alpha)/\beta)$
d. α **e.** $\mu = 201.95$, moda $= 150$, $\tilde{\mu} = 182.99$

111. **a.** μ **b.** No **c.** 0 **d.** $(\alpha - 1)\beta$ **e.** $\nu - 2$

113. **b.** $p(1 - \exp(-\lambda_1 x)) + (1 - p)(1 - \exp(-\lambda_2 x))$ para $x \ge 0$ **c.** $p/\lambda_1 + (1 - p)/\lambda_2$
d. $V(X) = 2p/\lambda_1^2 + 2(1 - p)/\lambda_2^2 - \mu^2$
e. 1, $CV > 1$ **f.** $CV < 1$

115. **a.** Lognormal **b.** 1 **c.** 2.72, .0185

119. **a.** Exponencial con $\lambda = 1$
c. Gamma con parámetros α y $c\beta$

121. **a.** $(1/365)^3$ **b.** $(1/365)^2$ **c.** .000002145

123. **b.** Sea $u_1, u_2, u_3, \ldots$ una sucesión de observaciones de una distribución Unif[0, 1] (una sucesión de números aleatorios). Entonces con $x_i = (-.1)\ln(1 - u_i)$ las x_i son observaciones de una distribución exponencial con $\lambda = 10$.

125. $g(E(X)) \le E(g(X))$

127. **a.** 710, 84.423, .684 **b.** .376

Capítulo 5

1. **a.** .20 **b.** .42 **c.** Al menos una manguera está en uso en cada bomba; .70. **d.** $p_X(x) = .16, .34, .50$ para $x = 0, 1, 2$, respectivamente; $p_Y(y) = .24, .38, .38$ para $y = 0, 1, 2$, respectivamente; .50 **e.** No; $p(0, 0) \ne p_X(0) \cdot p_Y(0)$

3. **a.** .15 **b.** .40 **c.** .22 **d.** .17, .46

5. **a.** .054 **b.** .00018

7. **a.** .030 **b.** .120 **c.** .300 **d.** .380 **e.** Sí

9. **a.** 3/380,000 **b.** .3024 **c.** .3593
d. $10Kx^2 + .05$ para $20 \le x \le 30$ **e.** No

11. **a.** $e^{-\mu_1 - \mu_2} \cdot \mu_1^x \cdot \mu_2^y/x!y!$ **b.** $e^{-\mu_1 - \mu_2} \cdot [1 + \mu_1 + \mu_2]$
c. $e^{-(\mu_1 + \mu_2)} \cdot (\mu_1 + \mu_2)^m/m!$; Poisson $(\mu_1 + \mu_2)$

13. **a.** e^{-x-y} para $x \ge 0, y \ge 0$ **b.** .400 **c.** .594
d. .330

15. **a.** $F(y) = 1 - e^{-\lambda y} + (1 - e^{-\lambda y})^2 - (1 - e^{-\lambda y})^3$ para $y \ge 0$
b. $2/3\lambda$

17. **a.** .25 **b.** .318 **c.** .637
d. $f_X(x) = 2\sqrt{R^2 - x^2}/\pi R^2$ para $-R \le x \le R$; no

19. **a.** $K(x^2 + y^2)/(10Kx^2 + .05)$; $K(x^2 + y^2)/(10Ky^2 + .05)$
b. .556, .549 **c.** 25.37, 2.87

21. **a.** $f(x_1, x_2, x_3)/f_{X_1, X_2}(x_1, x_2)$ **b.** $f(x_1, x_2, x_3)/f_{X_1}(x_1)$

23. .15

25. L^2

27. .25 h

29. $-\frac{2}{3}$

31. **a.** $-.1082$ **b.** $-.0131$

37. **a.**

$\bar{x}$	25	32.5	40	45	52.5	65
$p(\bar{x})$	.04	.20	.25	.12	.30	.09

$E(\bar{X}) = \mu = 44.5$

b.

s^2	0	112.5	312.5	800
$p(s^2)$	.38	.20	.30	.12

$E(S^2) = 212.25 = \sigma^2$

39.

Proporción	0	.1	.2	.3	.4	.5
Probabilidad	.000	.000	.000	.001	.005	.027

Proporción	.6	.7	.8	.9	1.0
Probabilidad	.088	.201	.302	.269	.107

41. **a.**

$\bar{x}$	1	1.5	2	2.5	3	3.5	4
$p(\bar{x})$	.16	.24	.25	.20	.10	.04	.01

b. .85 **c.**

r	0	1	2	3
$p(r)$	.30	.40	.22	.08

47. **a.** .6826 **b.** .1056

49. **a.** .6026 **b.** .2981

51. .7720

53. **a.** .0062 **b.** 0

55. **a.** .9838 **b.** .8926

57. .9616

59. **a.** .9986, .9986 **b.** .9875, .6266
c. .8357 **d.** .9525, .0003

61. **a.** 3.5, 2.27, 1.51 **b.** 15.4, 75.94, 8.71

63. **a.** .695 **b.** $4.0675 > 2.6775$

65. **a.** .9232 **b.** .9660

67. .1588

69. **a.** 2400 **b.** 1205; independencia **c.** 2400, 41.77

71. **a.** 158, 430.25 **b.** .9788

73. **a.** Aproximadamente normal con media = 105, DE = 1.2649; aproximadamente normal con media = 100, DE = 1.0142
b. aproximadamente normal con media = 5, DE = 1.6213
c. .0068 **d.** .0010, sí

75. **a.** .2, .5, .3 para $x = 12, 15, 20$; .10, .35, .55 para $y = 12, 15, 20$
b. .25 **c.** No **d.** 33.35 **e.** 3.85

77. **a.** 3/81,250 **b.** $f_X(x) = k(250x - 10x^2)$ para $0 \le x \le 20$ e $= k(450x - 30x^2 + \frac{1}{2}x^3)$ para $20 < x \le 30$; $f_Y(y)$ resulta de sustituir y por x en $f_X(x)$. No son independientes.
c. .355 **d.** 25.969 **e.** 204.6154, $-.894$ **f.** 7.66

79. ≈ 1

81. **a.** 400 min **b.** 70

83. 97

85. .9973

89. b, c. Chi cuadrada con $\nu = n$.

91. **a.** $\sigma_W^2/(\sigma_W^2 + \sigma_E^2)$ **b.** .9999

93. 26, 1.64

95. **a.** .6 **b.** $U = \rho X + \sqrt{1 - \rho^2}\, Y$

Capítulo 6

1. **a.** 8.14, $\bar{X}$ **b.** .77, $\tilde{X}$ **c.** 1.66, S
d. .148 **e.** .204, $S/\bar{X}$

3. **a.** 1.348, $\bar{X}$ **b.** 1.348, $\bar{X}$ **c.** 1.781, $\bar{X} + 1.28S$
d. .6736 **e.** .0846

5. $N\bar{x} = 1{,}703{,}000$; $T - N\bar{d} = 1{,}591{,}300$; $T \cdot (\bar{x}/\bar{y}) = 1{,}601{,}438.281$

7. **a.** 120.6 **b.** 1,206,000 **c.** .80 **d.** 120.0

9. **a.** 2.11 **b.** .119

11. **b.** $\left[\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right]^{1/2}$ **c.** Use $\hat{p}_i = x_i/n_i$ y $\hat{q}_i = 1 - \hat{p}_i$ en lugar de p_i y q_i en el inciso (b) para $i = 1, 2$.
d. $-.245$ **e.** .041

15. **a.** $\hat{\theta} = \sum X_i^2/2n$ **b.** 74.505

17. **b.** .444

19. **a.** $\hat{p} = 2\hat{\lambda} - .30 = .20$ **b.** $\hat{p} = (100\hat{\lambda} - 9)/70$

21. **b.** $\hat{\alpha} = 5$, $\hat{\beta} = 28.0/\Gamma(1.2)$

23. $\hat{\mu}_1 = \bar{x}$, $\hat{\mu}_2 = \bar{y}$, estimación de $(\mu_1 - \mu_2)$ es $\bar{x} - \bar{y}$.

25. **a.** 384.4, 18.86 **b.** 415.42

29. **a.** $\hat{\theta} = \text{mín}\,(X_i)$, $\hat{\lambda} = n/\sum[X_i - \text{mín}\,(X_i)]$
b. .64, .202

33. Con x_i = tiempo entre nacimiento $i - 1$ y nacimiento i, $\hat{\lambda} = 6/\sum_{i=1}^{6} i x_i = .0436$.

35. 29.5

37. 1.0132

Capítulo 7

1. a. 99.5% **b.** 85% **c.** 2.96 **d.** 1.15

3. a. Más angosto **b.** No **c.** No **d.** No

5. a. (4.52, 5.18) **b.** (4.12, 5.00) **c.** .55 **d.** 94

7. Por un factor de 4; el ancho se reduce en un factor de 5.

9. a. $(\bar{x} - 1.645\sigma/\sqrt{n}, \infty)$; $(4.57, \infty)$
b. $(\bar{x} - z_\alpha \cdot \sigma/\sqrt{n}, \infty)$ **c.** $(-\infty, \bar{x} + z_\alpha \cdot \sigma/\sqrt{n})$; $(-\infty, 59.7)$

11. 950, .8714

13. a. (608.58, 699.74) **b.** 189

15. a. 80% **b.** 98% **c.** 75%

17. 134.53

19. (.513, .615)

21. $p < .273$ con 95% de confianza; sí

23. a. $p > .044$ con 95% de confianza
b. Si la misma fórmula se utiliza en la muestra tras muestra, en el largo plazo el valor real de p superará el 95% de los límites inferiores calculados.

25. a. 381 **b.** 339

29. a. 2.228 **b.** 2.086 **c.** 2.845 **d.** 2.680
e. 2.485 **f.** 2.571

31. a. 1.812 **b.** 1.753 **c.** 2.602 **d.** 3.747
e. 2.1716 (de Minitab) **f.** Alrededor de 2.43

33. a. Cantidad razonable de simetría, sin valores atípicos
b. Sí (con base en una gráfica normal de probabilidad)
c. (430.5, 446.1), sí, no

35. a. 95% de intervalo de confianza: (23.1, 26.9)
b. 95% IP: (17.2, 32.8), casi 4 veces el ancho

37. a. (.888, .964) **b.** (.752, 1.100) **c.** (.634, 1.218)

39. a. Sí **b.** (6.45, 98.01) **c.** (18.63, 85.83)

41. Todos 70%; (c) porque es más corto

43. a. 18.307 **b.** 3.940 **c.** .95 **d.** .10

45. (3.6, 8.1); no

47. a. 95% de intervalo de confianza: (6.702, 9.456)
b. (.166, .410)

49. a. Parece haber una ligera asimetría positiva en la mitad central de la muestra, pero el bigote inferior es mucho más largo que el bigote superior. La magnitud de la variabilidad es más bien importante, aun cuando no haya resultados atípicos.
b. Sí. La figura de puntos en una gráfica de probabilidad normal es razonablemente lineal.
c. (33.53, 43.79)

51. a. (.539, .581) **b.** 2398 **c.** No—97.5%

53. $(-.84, -.16)$

55. 246

57. $(2t_r/\chi^2_{1-\alpha/2,\,2r},\ 2t_r/\chi^2_{\alpha/2,\,2r}) = (65.3, 232.5)$

59. a. $(\text{máx}\,(x_i)/(1 - \alpha/2)^{1/n},\ \ \text{máx}\,(x_i)/(\alpha/2)^{1/n})$
b. $(\text{máx}\,(x_i),\ \text{máx}\,(x_i)/\alpha^{1/n})$ **c.** (b); (4.2, 7.65)

61. (73.6, 78.8) contra (75.1, 79.6)

Capítulo 8

1. a. Sí **b.** No **c.** No
d. Sí **e.** No **f.** Sí

5. H_0: $\sigma = 0.5$ contra H_a: $\sigma < 0.5$. I: concluya que la variabilidad en grosor es satisfactoria cuando no lo es. II: concluya que la variabilidad en grosor no es satisfactoria cuando de hecho lo es.

7. I: concluyendo que la planta no se apega cuando sí se cumple; II: concluyendo que la planta se apega cuando de hecho no se cumple.

9. a. R_1 **b.** I: juzgando que una de las dos compañías es favorecida sobre la otra cuando ese no es el caso; II: juzgando que ninguna compañía es favorecida sobre la otra cuando de hecho las dos son realmente preferidas. **c.** .044
d. $\beta(.3) = \beta(.7) = .488$, $\beta(.4) = \beta(.6) = .845$
e. Rechazar H_0 a favor de H_a.

11. a. H_0: $\mu = 10$ contra H_a: $\mu \neq 10$ **b.** .01
c. .5319, .0078 **d.** 2.58
e. 10.1032 se sustituye con 10.124, y 9.8968 se sustituye con 9.876. **f.** $\bar{x} = 10.020$, de modo que H_0 no debe ser rechazada.
g. $z \geq 2.58$ o ≤ -2.58

13. b. .0004, 0, menor que .01

15. a. .0301 **b.** .003 **c.** .004

17. a. $z = 2.56 \geq 2.33$, y rechazar H_0. **b.** .8413 **c.** 143
d. .0052

19. a. $z = -2.27$, y no rechazar H_0. **b.** .2266 **c.** 22

21. a. $t_{.025,12} = 2.179 > 1.6$, y no rechazar H_0: $\mu = .5$.
b. $-1.6 > -2.179$, y no rechazar H_0.
c. No rechazar H_0.
d. Rechazar H_0 a favor de H_a: $\mu \neq .5$.

23. $t = 2.24 \geq 1.708$ y H_0 debe rechazarse. La información no sugiere una contradicción de creencia anterior.

25. **a.** $z = -3.33 \leq -2.58$ y rechazar H_0.
b. .1056 **c.** 217

27. **a.** $\bar{x} = .750$, $\tilde{x} = .640$, $s = .3025$, $f_s = .480$. Una gráfica de caja muestra importante asimetría positiva; no hay valores atípicos.
b. No. Una gráfica de probabilidad normal muestra curvatura importante. No, porque n es grande.
c. $z = -5.79$; rechazar H_0 a cualquier nivel razonable de significación; sí. **d.** .821

29. **a.** No, ya que $1.19 < 1.796$ **b.** .30 (con el software)

31. No, ya que $1.04 < 2.132$

33. **a.** No, ya que $2.44 < 2.539$ **b.** Sí, tipo II
c. .66 (con el software)

37. **a.** No, ya que $1.28 < 1.645$
b. I: dice que más de 20% son obesos cuando este no es el caso; II: concluye que 20% son obesos cuando el porcentaje real excede 20%.
c. .121

39. $z = 3.67 \geq 2.58$, y rechazar H_0: $p = .40$. No.

41. **a.** $z = -1.0$, así que no hay suficiente evidencia para concluir que $p < .25$; Por tanto se usan tapones de rosca.
b. I: No usar tapones de rosca cuando se justifica su uso;
II: Usar tapones de rosca cuando no se justifica su uso.

43. **a.** $z = 3.07 \geq 2.58$, rechazar H_0 y la premisa de la compañía.
b. .0332

45. No, no, sí. $R = \{5, 6, \ldots, 24, 25\}$, $\alpha = .098$, $\beta = .090$

47. **a.** Rechazar H_0. **b.** Rechazar H_0. **c.** No rechazar H_0.
d. Rechazar H_0. (una llamada cercana) **e.** No rechazar H_0.

49. **a.** .0778 **b.** .1841 **c.** .0250
d. .0066 **e.** .5438

51. **a.** .040 **b.** .018 **c.** .130 **d.** .653
e. $< .005$ **f.** $\approx .000$

53. Valor $P > \alpha$, y no rechazar H_0; no hay diferencia aparente.

55. Valor $P < .0004 < .01$, por lo que H_0: $\mu = 5$ podría ser rechazada a favor de H_a: $\mu \neq 5$.

57. No, ya que el valor $P = .2266$

59. **a.** Sí **b.** El valor P excede ligeramente .10, así que H_0: $\mu = 100$ no debería rechazarse y podría usarse el concreto.

61. $t \approx 1.9$, de modo que el valor $P \approx .116$. Por lo tanto, H_0 no debe ser rechazada.

63. **a.** .8980, .1049, .0014 **b.** Valor $P \approx 0$. Sí. **c.** No

65. $z = -3.12 \leq -1.96$, de modo que H_0 debe ser rechazada.

67. **a.** H_0: $\mu = .85$ contra H_a: $\mu \neq .85$
b. H_0 no puede ser rechazada por ninguna α.

69. **a.** No, porque el valor $P = .02 > .01$; sí, porque 45.31 excede por mucho 20, pero n es muy pequeña.
b. $\beta = .3$ (con software)

71. **a.** No; no
b. No, porque $z = .44$ y valor $P = .33 > .10$.

73. **a.** Aproximadamente .6; aproximadamente .2 (de la tabla A.17 del apéndice) **b.** $n = 28$

75. **a.** $z = 1.64 < 1.96$, so H_0 de modo que H_0 no puede ser rechazada; tipo II
b. .10. Sí.

77. Sí. $z = -3.32 \leq -3.08$, de modo que H_0 debe rechazarse.

79. No, porque $z = 1.33 < 2.05$.

81. Sí, porque $z = 4.4$ y el valor $P = 0 < .05$

83. **a.** $.01 <$ valor $P < .025$, de modo que no se rechace H_0; no hay contradicción

85. **a.** Para H_a: $\mu < \mu_0$, rechazar H_0 si $2\sum x_i/\mu_0 \leq \chi^2_{1-\alpha, 2n}$
b. Valor estadístico de prueba $= 19.65 > 8.260$, y no rechazar H_0.

87. **a.** Sí, $\alpha = .002$

Capítulo 9

1. **a.** $-.4$ h; no depende **b.** .0724, .2691 **c.** No

3. $z = 1.76 < 2.33$, de modo que no rechace H_0.

5. **a.** $z = -2.90$, y rechace H_0. **b.** .0019
c. .8212 **d.** 66

7. No, ya que el valor de P para la prueba de 2 colas de .0602.

9. **a.** 6.2; sí **b.** $z = 1.14$, valor $P \approx .25$, no
c. No **d.** Un intervalo de confianza de 95% es (10.0, 21.8).

11. Un intervalo de confianza de 95% es (.99, 2.41).

13. 50

15. **b.** Aumenta.

17. **a.** 17 **b.** 21 **c.** 18 **d.** 26

19. $t = -1.20 > -t_{.01,9} = -2.764$, no rechazar H_0.

21. Puesto que no es verdad que; $-2.46 \leq -2.602$, no rechazamos H_0. No se tiene suficiente evidencia.

23. **b.** No **c.** $t = -.38 > -t_{\alpha/2, 10}$ para cualquier α razonable, entonces no rechazar H_0 (valor $P \approx .7$).

25. (.3, 6.1), sí, sí

27. **a.** Intervalo de confianza de 99%: (.33, .71) **b.** Intervalo de confianza de 99%: $(-.07, .41)$, así que 0 es un posible valor de la diferencia.

29. $t = -2.10$, gl $= 25$, valor $P = .023$ d. Al nivel de significancia .05, concluiríamos que la cola resulta en un promedio de resistencia más alto, pero no al nivel de significancia de .01.

31. **a.** Centros prácticamente idénticos, sustancialmente más variabilidad en observaciones de rango medio que en observaciones de rango más alto
b. $(-7.9, 9.6)$, con base en 23 grados de libertad; no

Glosario de símbolos y abreviaturas

Símbolo/ Abreviatura	Página	Descripción
n	13	tamaño de muestra
x	13	variable en la que se hacen las observaciones
$x_1, x_2, \ldots, x_n$	13	observaciones muestrales en x
$\sum_{i=1}^{n} x_i$	28	suma de $x_1, x_2, \ldots, x_n$
$\bar{x}$	28	media muestral
μ	29	media poblacional
N	29	tamaño de la población cuando la población es finita
$\tilde{x}$	30	mediana muestral
$\tilde{\mu}$	31	mediana poblacional
$\bar{x}_{tr}$	32	media recortada
x/n	33	proporción de la muestra
s^2	36	varianza de la muestra
s	36	desviación estándar de la muestra
σ^2, σ	37	varianza poblacional y desviación estándar
$n - 1$	38	grados de libertad para una sola muestra
S_{xx}	38	suma del cuadrado de las desviaciones de la media muestral
f_s	39	dispersión de los cuartos muestrales
$\mathcal{S}$	51	espacio muestral de un experimento
$A, B, C_1, C_2, \ldots$	52	varios eventos
A'	53	complemento del evento A
$A \cup B$	53	unión de los eventos A y B
$A \cap B$	53	intersección de los eventos A y B
$\varnothing$	54	evento nulo (evento que no contiene resultados)
$P(A)$	56	probabilidad del evento A
N	61	número de resultados igualmente probables
$N(A)$	62	número de resultados en el evento A
n_1, n_2	65	número de maneras de seleccionar el 1er. (2o.) elemento de un par ordenado
$P_{k,n}$	67	número de permutaciones de tamaño k a partir de n entidades distintas
$\binom{n}{k}$	69	número de combinaciones de tamaño k a partir de n entidades distintas
$P(A \mid B)$	73	probabilidad condicional de A dado que B ocurre

Símbolo/ Abreviatura	Página	Descripción
rv	93	variable aleatoria
X	93	una variable aleatoria
$X(s)$	93	valor de la va X asociada con el resultado s
x	93	valor particular de va x
$p(x)$	96	distribución de probabilidad (función masa) de una va X discreta
pmf	97	función de masa de probabilidad
$p(x; \alpha)$	100	fmp con parámetro α
$F(x)$	101	función de distribución acumulada de una va
cdf	101, 144	función de distribución acumulada
$a-$	104	valor X más grande posible menor que a
$E(X), \mu_X, \mu$	108, 148	media o valor esperado de la va X
$E[h(X)]$	109, 149	valor esperado de la función $h(X)$
$V(X), \sigma_X^2, \sigma^2$	111, 150	varianza de la va X
σ_X, σ	111, 150	desviación estándar de la va X
S, F	114	éxito/fracaso en un ensayo simple de un experimento binomial
n	114	número de intentos en un experimento binomial
p	114, 125	probabilidad de éxito en un intento sencillo de un experimento binomial o un experimento binomial negativo
$X \sim \text{Bin}(n, p)$	116	la va X tiene una distribución binomial con parámetros n y p
$b(x; n, p)$	117	fmp binomial con parámetros n y p
$B(x; n, p)$	118	función de distribución acumulada de una va binomial
M	123	número de éxitos en una dicotomía poblacional de tamaño N
$h(x; n, M, N)$	123	fmp hipergeométrica con parámetros $n, M,$ y N
r	125	número de éxitos esperados en un experimento binomial negativo
$nb(x; r, p)$	125	fmp binomial negativa con parámetros r y p
μ	128	parámetro de una distribución de Poisson
$p(x; \mu)$	128	fmp de Poisson

Símbolo/ Abreviatura	Página	Descripción
$F(x; \mu)$	130	fda de Poisson
Δt	131	longitud de un intervalo corto de tiempo
$o(\Delta t)$	131	cantidad que se aproxima a 0 más rápido que Δt
α	131	parámetro de proporción de un proceso de Poisson
$\alpha(t)$	131	función de proporción de una tasa variable de un proceso de Poisson
fdp	139	función de densidad de probabilidad
$f(x)$	139	función de densidad de probabilidad de una va X continua
$f(x; A, B)$	140	fdp uniforme en el intervalo $[A, B]$
$F(x)$	144	función de distribución acumulada
$\eta(p)$	147	100p-ésimo percentil de una distribución continua
$\tilde{\mu}$	148	mediana de una distribución continua
$f(x; \mu, \sigma)$	152	fdp de una va normalmente distribuida
$N(\mu, \sigma^2)$	153	distribución normal con parámetros μ y σ^2
Z	153	va normal estándar
curva Z	154	curva normal estándar
$\Phi(z)$	154	fda de una va normal estándar
z_α	156	valor que captura un área de cola superior α bajo la curva z
λ	165	parámetro para una distribución exponencial
$f(x; \lambda)$	165	fdp exponencial
$\Gamma(\alpha)$	167	función gamma
$f(x; \alpha, \beta)$	168	fdp gamma con parámetros α y β
gl	170	grados de libertad
ν	170	número de gl para una distribución chi-cuadrada
$f(x; \alpha, \beta)$	172	fdp de Weibull con parámetros α y β
$f(x; \mu, \sigma)$	174	fdp lognormal con parámetros μ y σ
$f(x; \alpha, \beta, A, B)$	176	fdp beta con parámetros α, β, A, B
θ_1, θ_2	185	parámetros de ubicación y escala
$p(x, y)$	194	fmp conjunta de dos va discretas X y Y
$p_X(x), p_Y(y)$	195	fmp marginales de X y Y, respectivamente,
$f_X(x), f_Y(y)$	197	fdp marginales de X y Y, respectivamente
$p(x_1, \ldots, x_n)$	200	fmp conjunta de las n va $X_1, \ldots, X_n$
$f(x_1, \ldots, x_n)$	200	fdp conjunta de las n va $X_1, \ldots, X_n$
$f_{Y\|X}(y\|x)$	202	fdp condicional de Y dado que $X = x$
$p_{Y\|X}(y\|x)$	202	fmp condicional de Y dado que $X = x$
$E(Y\|X = x)$	203	valor esperado de Y dado que $X = x$
$E[h(X, Y)]$	206	valor esperado de la función $h(X, Y)$
$\mathrm{Cov}(X, Y)$	207	covarianza entre X y Y
$\mathrm{Corr}(X, Y), \rho_{X,Y}, \rho$	209	coeficiente de correlación para X y Y

Símbolo/ Abreviatura	Página	Descripción
$\overline{X}$	214	media muestral vista como una va
S^2	214	varianza muestral vista como una va
TLC	225	teorema del límite central
θ	240	símbolo genérico para un parámetro
$\hat{\theta}$	240	estimado puntual o estimador de θ
MVUE	247	estimador insesgado de mínima varianza (o estimado)
$\hat{\sigma}_{\hat{\theta},s}$	251	desviación estándar estimada de $\hat{\theta}$
$x_1^*, \ldots, x_n^*$	251	muestra bootstrap
$\hat{\theta}^*$	251	estimación de θ a partir de una muestra bootstrap
mle	259	estimación de máxima probabilidad (o estimador)
IC	270	intervalo de confianza
$100(1 - \alpha)\%$	272	intervalo de confianza para un IC
T	286	variable que tiene una distribución t
ν	286	parámetro de grados de libertad (gl) para una distribución t
t_ν	286	distribución t con ν gl
$t_{\alpha,\nu}$	287	valor que captura un área de cola superior α bajo la curva de densidad t_ν
IP	290	intervalo de predicción
$\chi^2_{\alpha,\nu}$	294	valor que captura un área de cola superior α bajo la curva de densidad chi-cuadrada con ν gl
H_0	301	hipótesis nula
H_a	301	hipótesis alternativa
α	304	probabilidad de un error tipo I
β	304	probabilidad de un error tipo II
μ_0	310	valor nulo en una prueba respecto a μ
Z	310	estadístico de prueba basado en una distribución normal estándar
μ'	313	valor alternativo de μ en el cálculo de β
$\beta(\mu')$	313	probabilidad de error tipo II cuando $\mu = \mu'$
T	316	estadístico de prueba basado en una distribución t
θ_0	323	valor nulo en una prueba respecto a θ
p_0	324	valor nulo en una prueba respecto a p
p'	325	valor alternativo de p en el cálculo de β
$\beta(p')$	325	probabilidad de error tipo II cuando $p = p'$
Ω_0, Ω_a	341	conjuntos disjuntos de valores de parámetros en una prueba de razón de verosimilitud
m, n	346	tamaños de muestra en problemas de dos muestras
Δ_0	347	valor nulo en una prueba respecto a $\mu_1 - \mu_2$
Δ'	350	valor alternativo de $\mu_1 - \mu_2$ en el cálculo de β

Símbolo/ Abreviatura	Página	Descripción
S_p^2	361	estimador agrupado de σ^2
D_i	366	diferencia $X_i - Y_i$ para el par (X_i, Y_i)
$\bar{d}$, s_D	368	diferencia de la media muestral, desviación estándar muestral- de las diferencias para parejas de datos
p	376	valor común de p_1 y p_2 cuando $p_1 = p_2$
F	382	va que tiene una distribución F
ν_1, ν_2	382	grado de libertad para el numerador y el denominador de una distribución F
F_{α,ν_1,ν_2}	382	valor que captura un área de cola superior α bajo una curva F con ν_1, ν_2 gl
ANOVA	391	análisis de varianza
I	392	número de poblaciones en un factor unitario ANOVA
J	394	tamaño común de muestra cuando tamaños de muestra son iguales
X_{ij}, x_{ij}	394	j-ésima observación en una muestra a partir de la i-ésima población
$\bar{X}_{i\cdot}$	394	media de las observaciones en una muestra a partir de la i-ésima población
$\bar{X}_{\cdot\cdot}$	394	media de todas las observaciones de un conjunto de datos
MSTr	395	media cuadrática de los tratamientos
MSE	395	media cuadrática del error
F	397	estadístico de prueba basado en una distribución F
$x_{i\cdot}$	397	total de observaciones en la i-ésima muestra
$x_{\cdot\cdot}$	397	gran total todas las observaciones
SST	398	suma total de cuadrados
SSTr	398	suma del cuadrado del tratamiento
SSE	398	suma del cuadrado de los errores
m, ν	402	parámetros para una distribución de rango estudentizado
$Q_{\alpha,m,\nu}$	402	valor que captura un área de cola superior bajo una curva de densidad asociada al rango estudentizado
μ	408	promedio de la población representada en un factor único ANOVA
$\alpha_1, \ldots, \alpha_I$	408	efectos de tratamiento en un factor único ANOVA
ϵ_{ij}	408	desviación de X_{ij} del valor de su media
$J_1, \ldots, J_I$	412	tamaño de muestra individual en un factor único ANOVA
n	412	número total de observaciones de un factor único ANOVA para un conjunto de datos
$A_1, \ldots, A_I$	414	efectos aleatorios en un factor único ANOVA

Símbolo/ Abreviatura	Página	Descripción
A, B	420	factores en un factor doble ANOVA
K_{ij}	420	número de observaciones cuando un factor A está en el nivel i y el factor B está en el nivel j
I, J	420	número de niveles de factores A y B, respectivamente
$\bar{X}_{i\cdot}$, $\bar{X}_{\cdot j}$	421	promedio de observaciones cuando A (B) está en el nivel i (j)
μ_{ij}	421	respuesta esperada cuando A está en el nivel i y B está en el nivel j
α_i, β_j	423	efecto de A (B) en el nivel i (j)
f_A, f_B	424	proporciones F para prueba de hipótesis sobre efectos de los factores
A_i, B_j	430	efecto de los factores en un modelo de efectos aleatorios
σ_A^2, σ_B^2	430	varianzas del efecto de los factores
K	433	tamaño de la muestra para cada par de niveles (i, j)
γ_{ij}	433	interacción entre A y B en los niveles i y j
A_i, B_j, G_{ij}	438	efectos en una combinación o en modelos de efectos aleatorios
α_i, β_j, δ_k	442	efectos principales en un ANOVA trifactorial
γ_{ij}^{AB}, γ_{ik}^{AC}, γ_{jk}^{BC}	442	interacciones de dos factores en un ANOVA trifactorial
γ_{ijk}	442	interacción de tres factores en un ANOVA trifactorial
I, J, K	442–443	número de niveles de A, B, C en un ANOVA trifactorial
β_1, β_0	472	pendiente intersección de una recta de regresión de una población
ε	472	desviación de Y a partir de su valor medio en una regresión lineal simple
σ^2	472	varianza de la desviación aleatoria ϵ
$\mu_{Y\cdot x^*}$	473	valor medio de Y cuando $x = x^*$
$\sigma_{Y\cdot x^*}^2$	473	varianza de Y cuando $x = x^*$
$\hat{\beta}_1$, $\hat{\beta}_0$	478	estimación de mínimos cuadrados de β_1 y β_0
S_{xy}	479	$\Sigma(x_i - \bar{x})(y_i - \bar{y})$
$\hat{y}_i$	481	valor pronosticado de y cuando $x = x_i$
SSE	483	suma de cuadrados del error (residuo)
SST	485	suma total de cuadrados S_{yy}
r^2	485	coeficiente de determinación
$S_{\hat{\beta}_1}$	493	desviación estándar estimada de $\hat{\beta}_1$
r, R	509	coeficiente de correlación muestral
e_i^*	524	residuo estandarizado
$\beta_i\,(i = 1, \ldots, k)$	543	coeficiente de x_i en una regresión polinomial
$\hat{\beta}_i$	544	estimación por mínimos cuadrados de β_i
R^2	546, 559	coeficiente de determinación múltiple
β_i^*	548	coeficiente en una regresión polinomial centrada

Símbolo/ Abreviatura	Página	Descripción
β_i	553	coeficiente de un predictor x_i de una regresión poblacional
$\hat{\beta}_i$	557	estimación por mínimos cuadrados de β_i
SSE_k, SSE_l	565	SSE para modelos y simplificados completos, respectivamente
Γ_k	579	estimación normalizada total esperada del error
C_k	579	estimación de Γ_k
h_{ii}	582	coeficiente de y_i en $\hat{y}_i$
$\chi^2_{\alpha,\nu}$	596	valor que captura un área de cola superior bajo la curva χ^2 con ν gl
χ^2	597	estadístico de prueba basado en una distribución chi-cuadrada
$p_{10}, \ldots, p_{k0}$	597	valores nulos de una prueba chi-cuadrada de una H_0 sencilla
$\pi_i(\theta)$	603	categoría de probabilidad como una función de los parámetros $\theta_1, \ldots, \theta_m$
I, J	613	número de poblaciones y categorías en cada población cuando se prueba la homogeneidad
I, J	613	número de categorías en cada uno de dos factores cuando se prueba la independencia
n_{ij}	614	número de individuos de una muestra de una población i que caen en la categoría j
$n_{.j}$	614	número total de individuos muestreados en la categoría j
p_{ij}	614	proporción de la población i en la categoría j
$\hat{e}_{ij}$	615	cantidad esperada estimada en la celda i, j
n_{ij}	617	número de muestras que caen en la categoría i del 1er. factor y la categoría j del 2o. factor
p_{ij}	617	proporción de población en la categoría i del 1er. factor y la categoría j del 2o. factor
S_+	627	estadístico de rango con signo
W	635	estadístico de rango con suma
K	645	estadístico de prueba Kruskal-Wallis
R_{ij}	645	rango de X_{ij} entre todas las N observaciones en el conjunto de datos
$\overline{R}_{i.}$	645	promedio de rangos para las observaciones en la muestra a partir de la población o tratamiento i
F_r	647	estadístico de prueba de Friedman
UCL	652	límite superior de control
LCL	652	límite inferior de control
C_p, C_{pk}	653–654	índices de capacidad del proceso
R	658	rango de la muestra
ARL	660	promedio a largo plazo
IQR	661	rango intercuartil
CUSUM	672	suma acumulada
OC	681	característica de operación
AQL	682	nivel aceptable de calidad
LTPD	682	porcentaje de tolerancia en un lote defectuoso
AOQ	685	calidad de salida promedio
AOQL	685	límite de la calidad de salida promedio
ATI	685	promedio del número total de inspecciones