

# Estadística para administración

CUARTA EDICIÓN

**David M. Levine**

**Timothy C. Krehbiel**

**Mark L. Berenson**



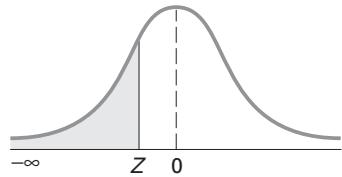
# MAPA PARA SELECCIONAR UN MÉTODO ESTADÍSTICO

## TIPO DE DATOS

Tipo de análisis	Numéricos	Categóricos
<b>Descripción de un grupo o diversos grupos</b>	<p>Arreglo ordenado, diagrama de tallo y hoja, distribución de frecuencias, distribución de frecuencia relativa, distribución de porcentajes, distribución de porcentajes acumulados, histograma, polígono, polígono de porcentaje acumulado (<b>Secciones 2.2 y 2.3</b>).</p> <p>Media, mediana, moda, cuartiles, media geométrica, rango, rango intercuartílico, desviación estándar, varianza, coeficiente de variación, gráfica de caja y bigote (<b>Secciones 3.1-3.3</b>).</p>	<p>Tabla de resumen, gráfica de barras, gráfica de pastel, diagrama de Pareto (<b>Sección 2.1</b>).</p>
<b>Inferencia acerca de un grupo</b>	<p>Estimación del intervalo de confianza para la media (<b>Secciones 8.1 y 8.2</b>).</p> <p>Prueba <i>Z</i> para la media (<b>Sección 9.2</b>).</p> <p>Prueba <i>t</i> para la media (<b>Sección 9.4</b>).</p>	<p>Estimación de intervalo de confianza para una proporción (<b>Sección 8.3</b>).</p> <p>Prueba <i>Z</i> de hipótesis para la proporción (<b>Sección 9.5</b>).</p>
<b>Comparación de dos grupos</b>	<p>Pruebas para la diferencia en las medias de dos poblaciones independientes (<b>Sección 10.1</b>).</p> <p>Prueba <i>t</i> apareada (<b>Sección 10.2</b>).</p> <p>Prueba <i>F</i> para la diferencia entre dos varianzas (<b>Sección 10.4</b>).</p>	<p>Prueba <i>Z</i> para la diferencia entre dos proporciones (<b>Sección 10.3</b>).</p> <p>Prueba de chi cuadrado para la diferencia entre dos proporciones (<b>Sección 11.1</b>).</p>
<b>Comparación de más de dos grupos</b>	<p>Análisis de varianza de una vía (<b>Sección 10.5</b>).</p>	<p>Prueba de chi cuadrado para las diferencias entre más de dos proporciones (<b>Sección 11.2</b>).</p>
<b>Ánalysis de la relación entre dos variables</b>	<p>Diagrama de dispersión, gráfica de series de tiempo (<b>Sección 2.5</b>).</p> <p>Covarianza, coeficiente de correlación (<b>Sección 3.4</b>).</p> <p>Regresión lineal simple (<b>Capítulo 12</b>).</p> <p>Prueba <i>t</i> de correlación (<b>Sección 12.7</b>).</p>	<p>Tabla de contingencia, gráfica de barras agrupadas (<b>Sección 2.4</b>).</p> <p>Prueba de chi cuadrado de independencia (<b>Sección 11.3</b>).</p>
<b>Ánalysis de la relación entre dos o más variables</b>	<p>Regresión múltiple (<b>Capítulo 13</b>).</p>	

## La distribución normal estandarizada acumulativa

Las entradas representan el área bajo la distribución normal estandarizada acumulativa desde  $-\infty$  hasta  $Z$

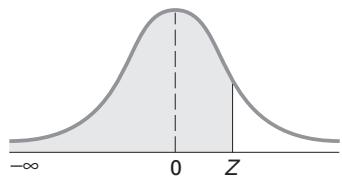


<b>Z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
-6.0	0.000000001									
-5.5	0.000000019									
-5.0	0.000000287									
-4.5	0.000003398									
-4.0	0.000031671									
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2388	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2482	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

(continúa)

### La distribución normal estandarizada acumulativa (continuación)

Las entradas representan el área bajo la distribución normal estandarizada acumulativa desde  $-\infty$  hasta  $Z$



# ¿HA PENSADO EN HACER ESTE LIBRO A SU *Medida*?

## EL PROGRAMA DE PRENTICE HALL JUSTO A TIEMPO EN CIENCIAS DE LA DECISIÓN

Usted puede combinar capítulos de este libro con capítulos de cualquiera de los títulos listados en la siguiente página para crear un libro de texto hecho a la medida de los requerimientos de su curso. Puede agregar su propio material o casos provenientes de nuestra extensa colección. Tómese unos minutos para revisar los libros de sus estantes, así como el contenido de nuestro sitio Web, con lo que podrá crear su libro de texto ideal.

### El programa Justo a Tiempo le ofrece:

- **Material de calidad para elegir:** Además de los libros listados, también tiene la opción de incluir cualquiera de los casos de Prentice Hall Custom Business Resources, que le da acceso a los casos (y notas de enseñanza en ocasiones disponibles) de Darden, Harvard, Ivey, NACRA y Thunderbird. La mayoría de los casos pueden revisarse en nuestro sitio Web.
- **Flexibilidad:** Elija únicamente el material que desea, ya sea de un solo título o de varios (más los casos) y organice la secuencia que requiere.
- **Apoyo instruccional:** Usted tiene acceso directo al texto específico del CD-ROM que acompaña a este libro, así como copias de escritorio de su libro JIT (Just-In-Time).
- **Materiales externos:** También cuenta con la opción de incluir hasta 20% del texto proveniente de libros y materiales ajenos a Prentice Hall Custom Business Resources.
- **Ahorros en costos:** Los estudiantes pagan sólo por el material que usted elija. El precio base es de USD \$6.00, más \$2.00 para material de casos, más \$.09 por página. El texto puede reunirse con otros libros de texto de Pearson para obtener un descuento del 10%. El material externo tiene un precio de \$.10 por página más las tarifas por permisos.
- **Calidad del producto final:** La portada del libro, así como la página del título incluirán su nombre, escuela, departamento, nombre de la materia y número de sección. El libro quedará impreso en blanco y negro, y perfectamente encuadrado. La tabla de contenidos estará hecha a la medida. El número de páginas seguirá una secuencia a lo largo del texto.

Visite nuestro sitio Web en [www.prenhall.com/custombusiness](http://www.prenhall.com/custombusiness) y cree su texto a la medida en nuestro “bookbuildsite” o descargue las formas para ordenar por Internet.

# EL PROGRAMA DE PRENTICE HALL

# Justo a Tiempo



**USTED PUEDE HACER SU LIBRO DE TEXTO A LA MEDIDA CON CAPÍTULOS DE CUALQUIERA  
DE LOS SIGUIENTES TÍTULOS DE PRENTICE HALL.\***

## ESTADÍSTICA PARA ADMINISTRACIÓN

- Berenson/Levine/Krehbiel, BASIC BUSINESS STATISTICS, 10/e
- Groebner/Shannon/Fry/Smith, BUSINESS STATISTICS, 6/e
- Levine/Stephan/Krehbiel/Berenson, STATISTICS FOR MANAGERS USING MICROSOFT EXCEL, 4/e
- Levine/Krehbiel/Berenson, BUSINESS STATISTICS: A FIRST COURSE, 4/e
- Newbold/Carlson/Thorne, STATISTICS FOR BUSINESS AND ECONOMICS, 5/e
- Shannon/Groebner/Fry/Smith, A COURSE IN BUSINESS STATISTICS, 3/e

## PRODUCCIÓN/ADMINISTRACIÓN DE OPERACIÓN

- Anupindi/Chopra/Deshmukh/Van Mieghem/Zemel, MANAGING BUSINESS PROCESS FLOWS
- Handfield/Nichols, Jr., SUPPLY CHAIN MANAGEMENT
- Hakssever/Render/Russell/Murdick, SERVICE MANAGEMENT AND OPERATIONS, 2/e
- Hanna/Newman, INTEGRATED OPERATIONS MANAGEMENT
- Heineke/Meile, GAMES AND EXERCISES IN OPERATIONS MANAGEMENT
- Heizer/Render, OPERATIONS MANAGEMENT, 7/e
- Krajewski/Ritzman, OPERATIONS MANAGEMENT, 7/e
- Latona/Nathan, CASES AND READINGS IN POM
- Russell/Taylor, OPERATIONS MANAGEMENT, 4/e
- Schmenner, PLANT AND SERVICE TOURS IN OPERATIONS MANAGEMENT, 5/e
- Nicholas, PROJECT MANAGEMENT, 2/e

## CIENCIA DE LA ADMINISTRACIÓN/MODELACIÓN DE HOJAS DE TRABAJO

- Eppen/Gould, INTRODUCTORY MANAGEMENT SCIENCE, 5/e
- Moore/Weatherford, DECISION MODELING WITH MICROSOFT EXCEL, 6/e
- Render/Stair/Hanna, QUANTITATIVE ANALYSIS FOR MANAGEMENT, 8/e
- Render/Stair/Balakrishnan, MANAGERIAL DECISION MODELING WITH SPREADSHEETS
- Render/Stair, CASES AND READINGS IN MANAGEMENT SCIENCE
- Taylor, INTRODUCTION TO MANAGEMENT SCIENCE, 8/e

Para mayor información, o para hablar con un representante del servicio a la medida,  
llame al 1-800-777-6872.

[www.prenhall.com/custombusiness](http://www.prenhall.com/custombusiness)

\* La selección de títulos del programa justo a tiempo (JIT) está sujeto a cambios



# ESTADÍSTICA PARA ADMINISTRACIÓN

CUARTA EDICIÓN

**DAVID M. LEVINE**

*Departamento de Estadística y Sistemas de Información Computacionales  
Zicklin School of Business, Baruch College, City University of New York*

**TIMOTHY C. KREHBIEL**

*Departamento de Ciencias de la Decisión y Administración de Sistemas de Información  
Richard T. Farmer School of Business, Miami University*

**MARK L. BERENSON**

*Departamento de Ciencias de la Información y Decisión  
School of Business, Montclair State University*

**Traducción:**

MARTHA L. GONZÁLEZ ACOSTA  
SERGIO A. DURÁN REYES  
*Traductores profesionales*

**Revisión técnica:**

DRA. OFELIA VIZCAÍNO DÍAZ  
Profesora del Departamento de Matemáticas  
Instituto Tecnológico y de Estudios Superiores  
de Monterrey, Campus Ciudad de México



MÉXICO • ARGENTINA • BRASIL • COLOMBIA • COSTA RICA • CHILE • ECUADOR  
ESPAÑA • GUATEMALA • PANAMÁ • PERÚ • PUERTO RICO • URUGUAY • VENEZUELA

Datos de catalogación bibliográfica	
<b>LEVINE, DAVID M., KREHBIEL, TIMOTHY C. Y MARK L. BERENSON</b>	
<b>Estadística para administración. Cuarta edición</b>	
PEARSON EDUCACIÓN, México, 2006	
ISBN: 970-26-0802-3	
Área: Matemáticas	
Formato: 21 x 27 cm	Páginas: 648

Authorized translation from the English language edition, entitled *Business statistics: a first course*, by David M. Levine, Timothy C. Krehbiel and Mark L. Berenson, published by Pearson Education, Inc., publishing as PRENTICE HALL, INC., Copyright ©2006. All rights reserved.

ISBN 013-153689-3

Traducción autorizada de la edición en idioma inglés, titulada *Business statistics: a first course* de David M. Levine, Timothy C. Krehbiel y Mark L. Berenson, publicada por Pearson Education, Inc., publicada como PRENTICE HALL INC., Copyright ©2006. Todos los derechos reservados.

Esta edición en español es la única autorizada.

#### Edición en español

Editor: Luis Miguel Cruz Castillo  
e-mail: luis.cruz@pearsoned.com  
Supervisor de desarrollo: Felipe Hernández Carrasco  
Supervisor de producción: Enrique Trejo Hernández

#### Edición en inglés

*Executive Editor:* Mark Pfaltzgraff  
*Editorial Director:* Jeff Shelstad  
*Managing Editor (Editorial):* Alana Bradley  
*Supplements Coordinator:* Kathryn Sheehan Madara  
*Senior Editorial Assistant:* Jane Avery  
*Product Development Manager:* Nancy Welcher  
*Executive Marketing Manager:* Debbie Clare  
*Marketing Assistant:* Joanna Sabella  
*Senior Managing Editor (Production):* Cynthia Regan  
*Senior Production Editor:* Anne Graydon  
*Permissions Supervisor:* Charles Morris  
*Production Manager:* Arnold Vila  
*Design Manager:* Maria Lange  
*Designer:* Steve Frim  
*Director, Image Resource Center:* Melinda Reo  
*Manager, Rights and Permissions:* Zina Arabia  
*Manager, Visual Research:* Beth Brenzel  
*Manager, Cover Visual Research & Permissions:* Karen Sanatar  
*Manager, Print Production:* Christy Mahon  
*Composition/Full-Service Project Management:* GGS Book Services, Atlantic Highlands  
*Printer/Binder:* Courier

CUARTA EDICIÓN, 2006

D.R. © 2006 por Pearson Educación de México, S.A. de C.V.  
Atlacomulco Núm. 500-5º piso  
Col. Industrial Atoto  
53519, Naucalpan de Juárez, Edo. de México  
E-mail: editorial.universidades@pearsoned.com

Cámara Nacional de la Industria Editorial Mexicana.  
Reg. Núm. 1031.

Prentice Hall es una marca registrada de Pearson Educación de México. S.A. de C.V.

Reservados todos los derechos. Ni la totalidad ni parte de esta publicación pueden reproducirse, registrarse o transmitirse, por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea electrónico, mecánico, fotoquímico, magnético o electroóptico, por fotocopia, grabación o cualquier otro, sin permiso previo por escrito del editor.

El préstamo, alquiler o cualquier otra forma de cesión de uso de este ejemplar requerirá también la autorización del editor o de sus representantes.



ISBN 970-26-0802-3

Impreso en México. *Printed in Mexico.*  
1 2 3 4 5 6 7 8 9 0 - 09 08 07 06

*Para nuestras esposas,  
Marilyn L., Patti K. y Rhoda B.*

*y para nuestros hijos,  
Sharyn, Ed, Rudy, Rhonda, Kathy y Lori*



# CONTENIDO BREVE

## Prefacio xxiii

1	INTRODUCCIÓN Y RECOLECCIÓN DE DATOS	1
2	PRESENTACIÓN DE DATOS EN TABLAS Y GRÁFICAS	21
3	MEDIDAS NUMÉRICAS DESCRIPTIVAS	71
4	PROBABILIDAD BÁSICA	121
5	ALGUNAS IMPORTANTES DISTRIBUCIONES DE PROBABILIDAD DISCRETA	153
6	LA DISTRIBUCIÓN NORMAL	177
7	DISTRIBUCIONES MUESTRALES	205
8	ESTIMACIÓN DE INTERVALOS DE CONFIANZA	237
9	FUNDAMENTOS DE LA PRUEBA DE HIPÓTESIS: PRUEBAS DE UNA MUESTRA	271
10	PRUEBAS DE DOS MUESTRAS Y ANOVA DE UNA VÍA	311
11	PRUEBAS DE CHI CUADRADA	377
12	REGRESIÓN LINEAL SIMPLE	409
13	REGRESIÓN MÚLTIPLE	465
14	APLICACIONES ESTADÍSTICAS EN ADMINISTRACIÓN DE LA CALIDAD Y PRODUCTIVIDAD	505

Apéndices A - G 537

Soluciones a los autoexámenes y respuestas a problemas pares seleccionados 589

Índice 615



# CONTENIDO

## Prefacio xxiii

### 1 INTRODUCCIÓN Y RECOLECCIÓN DE DATOS 1

Uso de la estadística: Good Tunes-Parte I	2
1.1 Conceptos básicos de la estadística	2
1.2 El crecimiento de la estadística y las tecnologías de la información	4
1.3 Cómo está organizado este texto	5
1.4 Recolección de datos	7
Identificación de las fuentes de datos	7
1.5 Tipos de datos	8
Resumen	9
Conceptos clave	10
Problemas de repaso	11
Introducción a los casos Web	11
A.1 Introducción del uso de los programas estadísticos	12
A1.1 Uso de Windows	12
A1.2 Introducción a Excel	13
A1.3 Introducción a Minitab	17
A1.4 ( <i>Tema CD-ROM</i> ) Introducción a SPSS	

### 2 PRESENTACIÓN DE DATOS EN TABLAS Y GRÁFICAS 21

Uso de la estadística: Comparación de los rendimientos de los fondos de inversión	22
2.1 Tablas y gráficas para datos categóricos	22
Tabla de resumen	22
Gráfica de barras	23
Gráfica de pastel	24
Diagrama de Pareto	25
2.2 Organización de los datos numéricos	29
Arreglo ordenado	30
Diagrama de tallo y hojas	30
2.3 Tablas y gráficas para datos numéricos	32
Distribución de frecuencias	32
Distribución de frecuencias relativas y distribución de porcentajes	34
Distribución acumulativa	35
Histograma	37
Polígono	38
Polígono de porcentaje acumulado (ojiva)	40

<b>2.4 Tablas y gráficas de datos bivariados</b>	<b>43</b>
Tabla de contingencia	43
Gráfica de barras agrupadas	45
<b>2.5 Diagramas de dispersión y series de tiempo</b>	<b>47</b>
Diagrama de dispersión	47
Series de tiempo	48
<b>2.6 Uso inadecuado de gráficas y consideraciones éticas</b>	<b>51</b>
Resumen	54
Conceptos clave	55
Problemas de repaso	55
Caso actual: Administración del <i>Springville Herald</i>	62
Caso Web	62
<b>A.2 Uso del software para tablas y gráficas</b>	<b>63</b>
A2.1 Excel	63
A2.2 Minitab	65
A2.3 ( <i>Tema CD-ROM</i> ) SPSS	

### **3 MEDIDAS NUMÉRICAS DESCRIPTIVAS 71**

#### **Uso de la estadística: Evaluación de los rendimientos de los fondos de inversión 72**

<b>3.1 Medidas de tendencia central, variación y forma</b>	<b>72</b>
La media	73
La mediana	75
La moda	76
Cuartiles	77
La media geométrica	79
Rango	80
Rango intercuartil	81
La varianza y la desviación estándar	82
Coeficiente de variación	85
Puntuaciones Z	86
Forma	88
Resultados de la estadística descriptiva en Excel	88
Exploraciones visuales: Exploración de la estadística descriptiva	89
Resultados de la estadística descriptiva en Minitab	89
<b>3.2 Medidas numéricas descriptivas de una población</b>	<b>94</b>
La media poblacional	94
Varianza y desviación estándar poblacionales	95
La regla empírica	96
La regla de Chebyshev	97
<b>3.3 Análisis exploratorio de datos</b>	<b>99</b>
Resumen de cinco números	99
Gráfica de caja y bigote	100
<b>3.4 La covarianza y el coeficiente de correlación</b>	<b>103</b>
La covarianza	103
Coeficiente de correlación	105

<b>3.5</b>	<b>Errores en las medidas numéricas descriptivas y consideraciones éticas</b>	<b>109</b>
	Aspectos éticos	110
	<b>Resumen</b>	<b>110</b>
	Conceptos clave	111
	Problemas de repaso	112
	Caso actual: Administración del <i>Springville Herald</i>	118
	Caso Web	118
	<b>A3 Uso de software para la estadística descriptiva</b>	<b>118</b>
	A3.1 Excel	118
	A3.2 Minitab	119
	A3.3 SPSS (tema del CD-ROM)	

## **4 PROBABILIDAD BÁSICA 121**

<b>Uso de la estadística: La empresa Consumer Electronics</b>	<b>122</b>
---	------------

<b>4.1</b>	<b>Conceptos básicos de probabilidad</b>	<b>122</b>
	Espacios muestrales y eventos	124
	Tablas de contingencia y diagramas de Venn	125
	Probabilidad simple (marginal)	125
	Probabilidad conjunta	127
	Regla general de la adición	128
<b>4.2</b>	<b>Probabilidad condicional</b>	<b>131</b>
	Cálculo de probabilidades condicionales	131
	Árboles de decisión	133
	Independencia estadística	134
	Reglas de multiplicación	136
	Probabilidad marginal usando la regla general de la multiplicación	137
<b>4.3</b>	<b>Teorema de bayes</b>	<b>139</b>
<b>4.4</b>	<b>Reglas de conteo</b>	<b>143</b>
<b>4.5</b>	<b>Consideraciones éticas y probabilidad</b>	<b>146</b>
	<b>Resumen</b>	<b>147</b>
	Conceptos clave	148
	Problemas de repaso	148
	Caso Web	150
	<b>A.4 Uso del software para la probabilidad básica</b>	<b>151</b>
	A4.1 Excel	151

## **5 ALGUNAS IMPORTANTES DISTRIBUCIONES DE PROBABILIDAD DISCRETA 153**

<b>Uso de la estadística: El sistema de información contable de la empresa de remodelaciones Saxon</b>	<b>154</b>
--	------------

<b>5.1</b>	<b>Distribución de probabilidad de una variable aleatoria discreta</b>	<b>154</b>
	Valor esperado de una variable aleatoria discreta	155
	Varianza y desviación estándar de una variable aleatoria discreta	156
<b>5.2</b>	<b>Distribución binomial</b>	<b>158</b>
<b>5.3</b>	<b>Distribución de Poisson</b>	<b>166</b>
	<b>Resumen</b>	<b>171</b>

Conceptos clave	171
Problemas de repaso	172
Caso actual: Administración del <i>Springville Herald</i>	174
<b>A.5 Uso de software para distribuciones de probabilidad discretas</b>	<b>175</b>
A5.1 Excel	175
A5.2 Minitab	176

## **6 LA DISTRIBUCIÓN NORMAL 177**

<b>Uso de la estadística: Tiempo de descarga para la página principal de un sitio Web</b>	<b>178</b>
<b>6.1 Distribuciones de probabilidad continua</b>	<b>178</b>
<b>6.2 La distribución normal</b>	<b>179</b>
<b>6.3 Evaluación de la normalidad</b>	<b>194</b>
Evaluación de las propiedades	194
Construcción de un plano de probabilidad normal	195
<b>Resumen</b>	<b>199</b>
<b>Conceptos clave</b>	<b>199</b>
<b>Problemas de repaso</b>	<b>199</b>
Caso actual: Administración del <i>Springville Herald</i>	201
<b>Caso Web</b>	<b>202</b>
<b>A.6 Uso del software con la distribución normal</b>	<b>202</b>
A6.1 Excel	202
A6.2 Minitab	202
A6.3 (Tema del CD-ROM) SPSS	

## **7 DISTRIBUCIONES MUESTRALES 205**

<b>Uso de la estadística: Proceso de empaquetado de cajas de cereal</b>	<b>206</b>
<b>7.1 Distribuciones muestrales</b>	<b>206</b>
<b>7.2 Distribución muestral de la media</b>	<b>207</b>
Propiedad de imparcialidad de la media muestral	207
Error estándar de la media	208
Muestreo de poblaciones con distribución normal	210
Muestreo de poblaciones sin distribución normal —Teorema del límite central	213
<b>7.3 Distribuciones muestrales de una proporción</b>	<b>217</b>
<b>7.4 Tipos de métodos de muestreo para encuestas</b>	<b>220</b>
Muestra aleatoria simple	221
Muestra sistemática	223
Muestra estratificada	224
Muestra de conglomerados	224
<b>7.5 Evaluación de las virtudes de una encuesta</b>	<b>226</b>
Errores de encuesta	226
Consideraciones éticas	228
<b>Resumen</b>	<b>229</b>
<b>Conceptos clave</b>	<b>230</b>
<b>Problemas de repaso</b>	<b>230</b>

<b>Caso actual: Administración del <i>Springville Herald</i></b>	<b>233</b>
<b>Caso Web</b>	<b>233</b>
<b>A.7 Uso de software para las distribuciones muestrales</b>	<b>234</b>
A7.1 Excel	234
A7.2 Minitab	234

## **8 ESTIMACIÓN DE INTERVALOS DE CONFIANZA 237**

<b>Uso de la estadística: Facturas de ventas auditadas en la empresa de remodelaciones Saxon</b>	<b>238</b>
<b>8.1 Estimación del intervalo de confianza para la media (<math>\sigma</math> conocida)</b>	<b>239</b>
<b>8.2 Estimación del intervalo de confianza para la media (<math>\sigma</math> desconocida)</b>	<b>243</b>
Distribución $t$ Student	243
Propiedades de la distribución $t$	244
El concepto de grados de libertad	245
El establecimiento del intervalo de confianza	246
<b>8.3 Estimación del intervalo de confianza de una proporción</b>	<b>250</b>
<b>8.4 Determinación del tamaño de la muestra</b>	<b>254</b>
Determinación del tamaño de la muestra para la media	254
Determinación del tamaño de la muestra para la proporción	256
<b>8.5 Estimación del intervalo de confianza y consideraciones éticas</b>	<b>260</b>
<b>Resumen</b>	<b>260</b>
<b>Conceptos clave</b>	<b>261</b>
<b>Problemas de repaso</b>	<b>261</b>
<b>Caso actual: Administración del <i>Springville Herald</i></b>	<b>265</b>
<b>Caso Web</b>	<b>266</b>
<b>A.8 Uso del software para los intervalos de confianza y la determinación del tamaño de la muestra</b>	<b>267</b>
A8.1 Excel	267
A8.2 Minitab	268
A8.3 (Tema del CD-ROM) SPSS	

## **9 FUNDAMENTOS DE LA PRUEBA DE HIPÓTESIS: PRUEBAS DE UNA MUESTRA 271**

<b>Uso de la estadística: Una visita más a la Oxford Cereal Company</b>	<b>272</b>
<b>9.1 Metodología de la prueba de hipótesis</b>	<b>272</b>
Hipótesis nula y alternativa	272
Valor crítico del estadístico de prueba	274
Regiones de rechazo y aceptación	274
Riesgos de la toma de decisiones al utilizar la metodología de la prueba de hipótesis	275
<b>9.2 Prueba Z de hipótesis para la media (<math>\sigma</math> conocida)</b>	<b>278</b>
Método del valor crítico para la prueba de hipótesis	278
Método del valor- $p$ para la prueba de hipótesis	281
Conexión entre la estimación del intervalo de confianza y la prueba de hipótesis	284

<b>9.3 Pruebas de una cola</b>	<b>286</b>
Método del valor crítico	286
Método del valor- <i>p</i>	287
<b>9.4 Prueba <i>t</i> de hipótesis para la media (<math>\sigma</math> desconocida)</b>	<b>290</b>
Método del valor crítico	291
Método del valor- <i>p</i>	293
Revisión de suposiciones	293
<b>9.5 Prueba Z de hipótesis para la proporción</b>	<b>297</b>
Método del valor crítico	298
Método del valor- <i>p</i>	299
<b>9.6 Posibles obstáculos en las pruebas de hipótesis y consideraciones éticas</b>	<b>301</b>
Resumen	304
Conceptos clave	305
Problemas de repaso	305
Caso actual: Administración del <i>Springville Herald</i>	308
Caso Web	308
<b>A.9 Uso de software para las pruebas de hipótesis con una muestra</b>	<b>309</b>
A9.1 Excel	309
A9.2 Minitab	309
A9.3 (Tema de CD-ROM) SPSS	

## **10 PRUEBAS DE DOS MUESTRAS Y ANOVA DE UNA VÍA** **311**

<b>Uso de la estadística: Comparación de las ventas de los exhibidores al final del pasillo y los exhibidores normales</b>	<b>312</b>
<b>10.1 Comparación de medias de dos poblaciones independientes</b>	<b>312</b>
Prueba <i>Z</i> para la diferencia entre dos medias	312
Prueba <i>t</i> de varianza conjunta para diferencias entre dos medias	313
Estimación del intervalo de confianza para la diferencia entre las medias de dos poblaciones independientes	318
Prueba <i>t</i> de varianza separada para la diferencia entre dos medias	318
<b>10.2 Comparación de medias de dos poblaciones relacionadas</b>	<b>322</b>
Prueba <i>t</i> apareada	324
Estimación del intervalo de confianza para la diferencia de la media	329
<b>10.3 Comparación de proporciones de dos poblaciones</b>	<b>332</b>
Prueba <i>Z</i> para la diferencia entre dos proporciones	332
Estimación del intervalo de confianza para la diferencia entre dos proporciones	336
<b>10.4 Prueba <i>F</i> para la diferencia entre dos varianzas</b>	<b>338</b>
Encontrar el valor crítico de la cola inferior	340
<b>Uso de la estadística: La empresa Perfect Parachute Company</b>	<b>346</b>
<b>10.5 Anova de una vía</b>	<b>346</b>
Prueba <i>F</i> para las diferencias entre más de dos medias	346
Comparaciones múltiples: el procedimiento Tukey-Kramer	353
Suposiciones de ANOVA	355
La prueba de la homogeneidad de la varianza de Levene	356

Resumen	361
Conceptos clave	363
Problemas de repaso	363
Caso actual: Administración del <i>Springville Herald</i>	369
Caso Web	370
<b>A.10 Uso del software para pruebas de dos muestras y ANOVA de una vía</b>	<b>371</b>
A10.1 Excel	371
A10.2 Minitab	373
A10.3 (Tema del CD-ROM) SPSS	

## **11 PRUEBAS DE CHI CUADRADA 377**

<b>Uso de la estadística: Satisfacción de los huéspedes en T.C. Resort Properties</b>	<b>378</b>
---	------------

<b>11.1 Prueba de chi cuadrada para la diferencia entre dos proporciones (muestras independientes)</b>	<b>378</b>
<b>11.2 Prueba de chi cuadrada para las diferencias entre más de dos proporciones</b>	<b>386</b>
Procedimiento de Marascuilo	389
<b>11.3 Prueba de independencia con chi cuadrada</b>	<b>393</b>
Resumen	399
Conceptos clave	401
Problemas de repaso	401
Caso actual: Administración del <i>Springville Herald</i>	405
Caso Web	406
<b>A.11 Uso de software para pruebas de chi cuadrada</b>	<b>407</b>
A11.1 Excel	407
A11.2 Minitab	407
A11.3 (Tema de CD-ROM) SPSS	

## **12 REGRESIÓN LINEAL SIMPLE 409**

<b>Uso de la estadística: Pronóstico de ventas para una tienda de ropa</b>	<b>410</b>
--	------------

<b>12.1 Tipos de modelos de regresión</b>	<b>410</b>
<b>12.2 Cómo determinar la ecuación de la regresión lineal simple</b>	<b>412</b>
El método de mínimos cuadrados	413
Exploraciones visuales: Explorando los coeficientes de la regresión lineal simple	416
Predicciones en el análisis de regresión: interpolación contra extrapolación	417
Cálculo de la intersección en Y, $b_0$ , y de la pendiente $b_1$	417
<b>12.3 Medidas de variación</b>	<b>421</b>
Cálculo de la suma de cuadrados	421
El coeficiente de determinación	424
Estimación del error estándar	426
<b>12.4 Suposiciones</b>	<b>428</b>
<b>12.5 Análisis residual</b>	<b>428</b>
Evaluación de las suposiciones	428
<b>12.6 Medición de la autocorrelación: estadístico de Durbin-Watson</b>	<b>433</b>
Gráfica residual para detectar la autocorrelación	433
El estadístico de Durbin-Watson	435

<b>12.7 Inferencias sobre la pendiente y el coeficiente de correlación</b>	<b>438</b>
Prueba $t$ para la pendiente	438
Prueba $F$ para la pendiente	440
Estimación del intervalo de confianza para la pendiente ( $\beta_1$ )	441
Prueba $t$ para el coeficiente de correlación	442
<b>12.8 Estimación de los valores de la media y predicción de los valores individuales</b>	<b>445</b>
La estimación del intervalo de confianza	445
El intervalo de predicción	447
<b>12.9 Dificultades de la regresión y consideraciones éticas</b>	<b>450</b>
Resumen	453
Conceptos clave	455
Problemas de repaso	455
Caso actual: Administración del <i>Springville Herald</i>	460
Caso Web	461
<b>A.12 Uso de software para la regresión lineal simple</b>	<b>462</b>
A12.1 Excel	462
A12.2 Minitab	464
A12.3 (Tema de CD-ROM) SPSS	

## **13 REGRESIÓN MÚLTIPLE 465**

<b>Uso de la estadística: Pronóstico de ventas de OmniPower</b>	<b>466</b>
<b>13.1 Desarrollo del modelo de regresión múltiple</b>	<b>466</b>
Interpretación de los coeficientes de regresión	467
Pronóstico de la variable dependiente $Y$	470
<b>13.2 <math>r^2, r^2</math> ajustada y prueba <math>F</math> global</b>	<b>472</b>
Coeficiente de determinación múltiple	472
Prueba de la significancia del modelo de regresión múltiple global	473
<b>13.3 Análisis residual para el modelo de regresión múltiple</b>	<b>476</b>
<b>13.4 Inferencias respecto a los coeficientes de regresión poblacionales</b>	<b>478</b>
Pruebas de hipótesis	478
Estimación del intervalo de confianza	480
<b>13.5 Uso de variables indicadoras y términos de interacción en los modelos de regresión</b>	<b>482</b>
Interacciones	484
<b>13.6 Modelo de regresión cuadrática</b>	<b>488</b>
Cómo encontrar los coeficientes de regresión y pronóstico de $Y$	489
Prueba de la significancia del modelo cuadrático	492
Prueba del efecto cuadrático	492
Resumen	496
Conceptos clave	498
Problemas de repaso	498
Caso actual: Administración del <i>Springville Herald</i>	501
Caso Web	501
<b>A.13 Uso de software para la regresión múltiple</b>	<b>502</b>
A13.1 Excel	502

A13.2 Minitab 503  
A13.3 (Tema de CD-ROM) SPSS

## 14 APLICACIONES ESTADÍSTICAS EN ADMINISTRACIÓN DE LA CALIDAD Y PRODUCTIVIDAD 505

- Uso de la estadística: Servicio de calidad en el hotel Beachcomber 506
- 14.1 Administración de calidad total 506
- 14.2 Administración Seis Sigma 509
- 14.3 La teoría de gráficas de control 509
- 14.4 Gráfica de control para la proporción de artículos disconformes: la gráfica  $p$  512
- 14.5 El experimento de la cuenta roja: comprendiendo el proceso de variabilidad 518
- 14.6 Gráficas de control para el rango y la media 520
- La gráfica  $R$  521
  - La gráfica  $\bar{X}$  523
- Resumen 527
- Conceptos clave 528
- Problemas de repaso 528
- Caso actual: Administración del *Springville Herald* 533
- A.14 Uso de software para las gráficas de control 535
- A14.1 Excel 535
  - A14.2 Minitab 535
  - A14.3 (Tema de CD-ROM) SPSS

## APÉNDICES 537

- A. Repaso de aritmética, álgebra y logaritmos 538
  - B. Notación de sumatoria 540
  - C. Símbolos estadísticos y alfabeto griego 543
  - D. Contenido del CD-ROM 544
  - E. Tablas 549
  - F. Uso de Excel con este libro 574
  - G. Guía del usuario de PHStat2 576
- Soluciones a los autoexámenes y repuestas a problemas pares seleccionados 589
- Índice 615



# PREFACIO

## Filosofía educacional

Durante muchos años de enseñar estadística a estudiantes de administración, continuamente hemos buscado formas para mejorar estos cursos. Nuestra activa participación en las series “Haciendo la estadística más efectiva en las escuelas y en la administración” en el Instituto de Ciencias de la Decisión y en las conferencias de la American Statistical Association, así como la oportunidad de atender a un diversificado grupo de estudiantes en las grandes universidades, ha moldeado nuestra visión de la enseñanza de estos cursos. A lo largo de los años, nuestra visión ha llegado a incluir estos principios clave:

1. Se necesita mostrar a los estudiantes la importancia de la estadística.
  - Los estudiantes necesitan un marco de referencia cuando aprenden estadística, especialmente cuando ésta no es su especialidad. Ese marco de referencia para los estudiantes de administración debe comprender las áreas funcionales de negocios, esto es, contabilidad, economía y finanzas, sistemas de información, administración y marketing. Cada tema estadístico necesita presentarse en un contexto aplicado relacionado por lo menos con una de estas áreas funcionales.
  - El enfoque de la enseñanza en cada tema debería partir de su aplicación a la administración, la interpretación de resultados, la presentación de suposiciones, la evaluación de las suposiciones y la discusión acerca de lo que debería hacerse si las suposiciones son infringidas.
2. Los estudiantes deben familiarizarse con el software que se usa en el mundo de los negocios.
  - Integrar las hojas de trabajo o el software estadístico en todos los aspectos de un curso de introducción a la estadística permite que el curso se enfoque en la interpretación de los resultados antes que en los cálculos.
  - Los cursos de introducción a la estadística en la administración deben reconocer que en los negocios el software de hojas de trabajo está disponible en el escritorio de toma de decisiones (en ocasiones también hay software estadístico).
3. Los estudiantes requieren de suficiente guía para usar el software.
  - Los libros de texto deben proporcionar suficientes instrucciones como para que los alumnos puedan usar de forma efectiva el software e integrarlo al estudio de la estadística, sin permitir que la instrucción del software domine el curso.
4. Los estudiantes requieren de suficiente práctica para comprender cómo se usa la estadística en la administración.
  - Es recomendable que los ejemplos en clase y los ejercicios de tarea incluyan datos actuales o reales tanto como sea posible.
  - Los alumnos deben trabajar con conjuntos de datos, tanto pequeños como grandes, y ser animados a ver más allá del análisis estadístico de los datos para interpretar los resultados en un contexto administrativo.

## Las innovaciones en esta edición

Esta cuarta edición de *Estadística para administración* ha mejorado en varias áreas importantes.

## Es más accesible para los estudiantes

- Cada capítulo en este texto ha pasado por una revisión mayor y ahora se utiliza un estilo de escritura más activo y conversacional que los estudiantes apreciarán. Las oraciones se han acortado y simplificado.
- El texto se enfoca ahora más en aquellos temas que se cubren de forma característica en un primer curso. La cobertura de la regresión múltiple se redujo y se eliminó el pronóstico de las series de tiempo.

- En el texto se incluyen ahora muchos más ejemplos de la vida cotidiana. Tales ejemplos incluyen aquellos acerca de las compras *on line* (capítulo 2), el tiempo para estar listo en la mañana (capítulo 3), y el tiempo de espera en un restaurante de comida rápida (capítulo 9).
- Hemos simplificado muchos problemas para que no contengan más de cuatro partes.
- Se incluyen fórmulas importantes al final de cada capítulo.
- Al final del texto se proporcionan las soluciones a las preguntas de los autoexámenes.
- Se incluye un mapa para seleccionar el método estadístico adecuado al inicio del texto para ayudar a los estudiantes a seleccionar la técnica apropiada y para realizar conexiones entre los temas.
- Se han agregado a este texto muchos ejemplos y ejercicios nuevos tomados de *The Wall Street Journal*, *USA Today* y *Consumer Reports*, así como de otras fuentes.
- En casi todos los capítulos se incluye al final un caso web. Al visitar sitios Web relacionados con las empresas e investigar los temas que surgen del escenario “Uso de la estadística” al inicio de cada capítulo, los estudiantes aprenden a identificar el mal uso que se le da a la información estadística. Los casos Web requieren que los estudiantes examinen cuidadosamente las afirmaciones y la información variada para descubrir qué datos son los más relevantes para el caso. Los alumnos entonces deberán determinar si los datos apoyan las conclusiones y afirmaciones. (En el Manual de soluciones para el instructor se incluyen algunos consejos para usar los casos Web.)

## Mayor instrucción de software

- Los apéndices de Excel al final de cada capítulo ahora explican cómo utilizar las hojas de trabajo estándar de Excel para realizar la mayoría de los análisis estadísticos. Instructores y estudiantes que desean evitar el uso de anexos, encontrarán de inmediato la utilidad de estas nuevas instrucciones. (Quienes elijan utilizar el apéndice referente a PHStat2 de Excel, encontrarán que se han colocado juntas todas las explicaciones de los comandos de PHStat2 en un nuevo apéndice G para una fácil referencia.)

A	B
1 Estimate for the Mean Sales Invoice Amount	
2	
3 Data	
4 Sample Standard Deviation	28.95
5 Sample Mean	110.27
6 Sample Size	100
7 Confidence Level	95%
8	
9 Intermediate Calculations	
10 Standard Error of the Mean	2.8950
11 Degrees of Freedom	99
12 t Value	1.9842
13 Interval Half Width	5.7443
14	
15 Confidence Interval	
16 Interval Lower Limit	104.53
17 Interval Upper Limit	116.01

=B4/SQRT(B6)  
 =B6 - 1  
 =TINV(1-B7,B11)  
 =B12 \* B10  
 =B5 - B13  
 =B5 + B13

- Muchas de las hojas de trabajo estándar de Excel examinadas en los apéndices Excel están incluidas como ilustraciones dentro de los capítulos. Cada ilustración (véase el ejemplo de arriba) incluye un listado de todas las fórmulas de celda contenidas en la hoja de trabajo. (Los usuarios de PHStat2 también encontrarán informativas estas ilustraciones ya que son consistentes con las hojas de trabajo que el PHStat2 genera.)
- Con este texto se incluye gratis una versión actualizada de PHStat2: PHStat2 versión 2.5, la versión más nueva de agregados de Prentice Hall para Excel. Esta versión actualizada incluye elementos como la regresión múltiple con variables independientes en columnas no contiguas, diagramas de tallo y hojas y gráficas de caja y bigote mejorados, la prueba Z para la diferencia de dos medias, la prueba de Levene para la homogeneidad de la varianza y el procedimiento de Marascuilo de múltiples comparaciones para proporciones. (Encontrará apoyo para el uso de PHStat2, incluyendo actualizaciones gratis cuando estén disponibles, en [www.prenhall.com/phstat](http://www.prenhall.com/phstat).)

- Uso Minitab, versión 14, la más reciente del software estadístico de Minitab. Todos los resultados y todos los apéndices de Minitab en este texto provienen de esa versión, la última del software estadístico de Minitab.

## Reorganización de los capítulos acerca de la prueba de hipótesis

- Todas las pruebas que implican la distribución normal y la distribución *t* se estudian en los capítulos 9 y 10 *antes* de la cobertura de la prueba *F*.
- El análisis de varianza se estudia en el capítulo 10.
- Todas las pruebas de chi cuadrada se ven en el capítulo 11.

## Cambios por capítulos en la cuarta edición

Cada capítulo tiene una nueva página inicial que indica las secciones y apartados que lo componen.

- *Capítulo 1.* Se rescribieron las secciones 1.1, 1.2 y 1.3. Los apartados que tratan acerca del muestreo de encuestas se cambiaron al capítulo 7.
- *Capítulo 2.* Contiene nuevos datos acerca de los rendimientos de los fondos de inversión para el periodo comprendido entre 1999 y 2003. Las gráficas para las variables categóricas se presentan antes que las gráficas para variables numéricas. Todas las gráficas para una variable se exponen antes que las gráficas para dos variables. Se agregan ejemplos dentro del capítulo concernientes a las ventas *on line* y al costo de los alimentos en restaurantes, además de los referentes a los fondos de inversión.
- *Capítulo 3.* Cuenta con un nuevo conjunto de datos concernientes a los rendimientos entre 1999 y 2003 de los fondos de inversión. Los ejemplos dentro del capítulo se refieren al tiempo que se requiere para prepararse en la mañana así como a los rendimientos de fondos de inversión, se incluyen ahora puntuaciones *Z* para detectar valores extremos incluidos. La covarianza de la muestra se presenta ahora junto con el coeficiente de correlación.
- *Capítulo 4.* Ahora incluye el teorema de Bayes y las reglas de conteo.
- *Capítulo 5.* Abarca ahora la distribución de Poisson. La distribución normal se pasó al capítulo 6.
- *Capítulo 6.* Se dedica por completo a la distribución normal e incluye una sección simplificada de la gráfica de probabilidad normal.
- *Capítulo 7.* Incluye las distribuciones muestrales y los métodos de tipos y muestreo de encuesta; se explica el valor de la encuesta.
- *Capítulo 8.* Incluye la estimación del intervalo de confianza y cómo determinar el tamaño de la muestra.
- *Capítulo 9.* Utiliza un método más simple de seis pasos para realizar pruebas de hipótesis usando el enfoque del valor crítico y un método claro de cinco pasos para realizar la prueba de hipótesis utilizando el enfoque del valor-*p*.
- *Capítulo 10.* Está reorganizado para que las pruebas de dos muestras para la media y la proporción precedan a la prueba *F* de la diferencia entre varianzas. El capítulo también incluye el ANOVA de una vía.
- *Capítulo 11.* Incluye únicamente las pruebas  $\chi^2$ .
- *Capítulo 12.* Ahora incluye cálculos para los coeficientes de regresión y suma de cuadrados en los ejemplos del capítulo.
- *Capítulo 13.* Ahora abarca  $r^2$ ,  $r^2$  ajustada y la prueba *F* global antes del análisis residual. El capítulo también incluye la regresión cuadrática.
- *Capítulo 14.* Se incluye la administración Seis Sigma.

## Características distintivas

Hemos dado continuidad a muchos elementos tradicionales de ediciones anteriores. Hemos resaltando algunas de las siguientes características:

- **Escenarios de negocios “Uso de la estadística”** Cada capítulo inicia con un ejemplo del “Uso de la estadística”, que muestra cómo la estadística se utiliza en la contabilidad, las finanzas, la administración o el marketing. Cada escenario se usa a lo largo del capítulo para ofrecer un contexto aplicado para los conceptos.

## USO DE LA ESTADÍSTICA



### Comparación de los rendimientos de los fondos de inversión

Entre las muchas opciones de inversión disponibles en la actualidad, una elección común para quienes piensan en su retiro son los fondos de inversión. Si usted decide invertir en fondos de inversión su cuenta de retiro, ¿qué haría para hacer una elección razonable de entre todos los fondos disponibles hoy?

Primero debería conocer las diferentes categorías de los fondos de inversión. Debería conocer las estrategias de los profesionales que administran los fondos. ¿Invierten en valores de alto riesgo o hacen elecciones más conservadoras? ¿El fondo se especializa en un determinado tamaño de compañía, uno cuya reserva principal totaliza un gran capital o uno de capital reducido? ¿Cobra el fondo comisiones por administración que reducen el porcentaje de utilidad del inversionista? Y, por supuesto, debería conocer qué tan bien el fondo ha manejado las inversiones en el pasado.

Todos éstos son datos que debe revisar cuando considere varias posibilidades de invertir en fondos de inversión. ¿Cómo “poner manos a la obra” con estos datos y explorarlos de manera exhaustiva?

- **Énfasis en el análisis de datos e interpretación de los resultados en computadora.** Creamos que el uso del software es una parte integral del aprendizaje de la estadística. Nuestro enfoque destaca el análisis de datos interpretando los resultados de Excel y Minitab, al mismo tiempo que reduce el énfasis en hacer cálculos. Por esa razón, hemos incluido más resultados de computadora y los hemos integrado al texto. Por ejemplo, en las tablas y gráficas del capítulo 2, el enfoque es en la interpretación, no en su construcción manual. En nuestra exposición de la prueba de hipótesis del capítulo 9 hasta el 11, se incluyen de forma extensa los resultados de computadora para que el énfasis se ponga en el método del valor-*p*. En nuestra explicación de la regresión lineal simple del capítulo 12, suponemos que se utiliza Excel o Minitab. Así que el énfasis está en la interpretación del resultado y no en los cálculos manuales.
- **Auxiliares pedagógicos.** Se incluyen en cada capítulo, con un estilo de escritura sencillo, cuadros de ecuaciones numeradas, una serie de ejemplos para reforzar los conceptos aprendidos, problemas divididos en Aprendizaje básico y Aplicación de conceptos, así como conceptos clave.
- **Apéndices al final del capítulo.** Usando Excel y la versión 14 de Minitab con ilustraciones, se incluyen instrucciones fáciles de seguir. Las instrucciones para PHStat2 se incluyen en el apéndice G. Los apéndices SPSS se incluyen en el CD-ROM que acompaña este texto.
- **Respuestas.** Se incluyen las respuestas para la mayoría de los ejercicios pares al final del libro.
- **PHStat2.** En el CD-ROM para estudiantes se incluye un programa complementario para Excel que incrementa sus capacidades estadísticas y ejecuta para usted el menú de selección de bajo nivel y las entradas de las tareas de hojas de trabajo asociadas con la realización de los análisis estadísticos en Excel. Al combinarlo con el Data Analysis ToolPak de Excel, se pueden ilustrar virtualmente todos los métodos estadísticos enseñados en un curso de introducción a la estadística utilizando Excel.
- **Estudios de caso y proyectos en equipo.** Se incluyen estudios de caso detallados al final de varios capítulos. El caso del *Springville Herald* se incluye al final de prácticamente todos los capítulos como tema integrador. Al final de muchos capítulos se incluye un Proyecto en equipo relacionado con los fondos de inversión como tema integrador.
- **Exploraciones visuales.** Se incluye gratuitamente con este texto un libro de trabajo de Excel, que permite a los estudiantes explorar de forma interactiva conceptos importantes en estadística descriptiva, probabilidad, distribución normal y análisis de regresión. Por ejemplo, en la estadística descriptiva, los estudiantes observan el efecto que tiene el cambio de datos en la media, mediana, cuartiles y desviación estándar. En las distribuciones muestrales, los estudiantes utilizan la simulación para explorar el efecto del tamaño de la muestra en la distribución muestral. Con la distribución normal, los estudiantes logran ver el efecto que tienen los cambios en la media y la desviación estándar sobre las áreas bajo la curva normal. En el análisis de regresión, los alumnos tienen la oportunidad de ajustar una línea y observar cómo cambia la pendiente y cómo la intersección afecta el ajuste. (Exploraciones visuales requiere la instalación del sistema de seguridad Medio de Excel.)

## Paquete complementario

El paquete complementario que acompaña este texto incluye el siguiente material:

- **Manual de soluciones para el instructor.** Este manual incluye recomendaciones de enseñanza para cada capítulo, detalles extra para la solución de problemas y muchas soluciones para Excel y Minitab.
- **Manual de soluciones para el estudiante.** Este manual proporciona soluciones detalladas para prácticamente todos los ejercicios pares.
- **Archivo de reactivos de examen.** El archivo de reactivos de examen contiene preguntas de cierto/falso, de opción múltiple, llenado y solución de problemas con base en las definiciones, conceptos e ideas desarrollados en cada capítulo del texto.
- **Software de exámenes TestGen.** Este banco de exámenes impreso está diseñado para utilizarse con el software generador de exámenes TestGen. Este paquete computarizado permite a los instructores diseñar, guardar y generar exámenes para el salón de clases. El programa permite a los maestros editar, aumentar o borrar las preguntas del banco de exámenes; editar las gráficas existentes y crear nuevas; analizar los resultados de los exámenes, y organizar una base de datos de las pruebas y de los resultados de los alumnos. Este software permite una mayor flexibilidad y es fácil de utilizar. Ofrece muchas opciones para organizar y presentar las pruebas, junto con una característica de buscar y seleccionar. El programa está disponible tanto en el CD-ROM del instructor como en el catálogo *on line* de Prentice Hall para su descarga.
- **Centro de recursos para el instructor.** Este centro de recursos para el instructor contiene los archivos electrónicos completos para el Manual de soluciones del instructor (en Word), el Archivo de reactivos de examen (en Word), el Archivo computarizado de reactivos de examen (Word), TestGen, y presentaciones en PowerPoint.
- **Herramientas administrativas para el curso y tareas**
- **Prentice Hall's OneKey.** Ofrece los mejores recursos de enseñanza y aprendizaje en un solo lugar. Todo lo que usted necesita para planear y administrar su curso lo encontrará en OneKey para *Estadística en la administración, Un primer curso, 4a. edición*, y es todo lo que sus alumnos necesitan para tener acceso a los materiales de su curso en cualquier momento y desde cualquier lugar. La compilación de recursos está convenientemente organizada siguiendo los capítulos del libro de texto e incluye: vínculos a exámenes, presentaciones en PowerPoint, archivos de datos, vínculos a los casos Web, descarga de PHStat2, Descarga de Exploraciones visuales, Manual de soluciones para el estudiante, así como recursos instruccionales adicionales.
- **WebCT y Pizarrón.** Con la instalación local de algún curso de sistema de administración, Prentice Hall brinda contenidos diseñados específicamente para este libro de texto con el fin de crear un juego completo para el curso, integrado estrechamente con las herramientas del sistema de administración del curso.
- **PH Grade Assist.** Este sistema *on line* de tareas y evaluación permite al instructor asignar problemas para que los estudiantes practiquen y resuelvan tareas o exámenes. Los problemas, tomados directamente del texto, son generados algorítmicamente, para que cada estudiante reciba un problema ligeramente diferente con una respuesta distinta. Esta característica permite al alumno realizar intentos múltiples para adquirir más práctica y mejorar su competencia. El PH Grade Assist califica los resultados y los traslada a las hojas de trabajo de Excel.
- **Sitio Web acompañante.** Este sitio contiene:
  - Una guía de estudio en línea con preguntas de cierto/falso, de opción múltiple y de ensayo, diseñadas para probar la comprensión del alumno en los temas del capítulo.
  - Archivos con presentaciones en PowerPoint con lo más importante del capítulo y sus correspondientes fórmulas.
  - Archivos de datos de los alumnos para los problemas del texto en Excel, Minitab y SPSS.
- **Versión Minitab para el estudiante.** Por un costo adicional razonable, se puede incluir una versión para el alumno de Minitab Versión 14. Por favor, póngase en contacto con su representante de ventas de Prentice Hall para información sobre cómo hacer un pedido.
- **Versión SPSS para el alumno.** Por un costo adicional razonable, se puede incluir una versión del SPSS 12 con este texto. Por favor, póngase en contacto con su representante de ventas de Prentice Hall para información sobre cómo hacer un pedido.
- **Sitio Web del texto.** Este texto tiene una página en la World Wide Web en [www.prenhall.com/levine](http://www.prenhall.com/levine). Este sitio ofrece muchos de los recursos tanto para maestros como para alumnos. **PHStat2** tiene una página en la World Wide Web en [www.prenhall.com/phstat](http://www.prenhall.com/phstat).

Es posible encontrar una página índice para el material de apoyo para todos los casos Web incluidos en este texto en [www.prenhall.com/Springville/Springvillecc.htm](http://www.prenhall.com/Springville/Springvillecc.htm).

## Agradecimientos

Estamos especialmente agradecidos con muchas organizaciones y empresas que nos permitieron utilizar sus datos para desarrollar problemas y ejemplos a lo largo del texto. Quisiéramos agradecer a *The New York Times*, Consumer Union (editores de *Consumer Reports*), Mergent's Investor Service (editores de *Mergent's Handbook of Common Stocks*) y CEEPress.

Además, quisiéramos dar las gracias a Biometrika Trustees, American Cyanimid Company, Rand Corporation y la American Society for Testing and Materials (por su amable permiso para publicar varias tablas en el apéndice E), así como a la American Statistical Association (por su permiso para publicar diagramas de *American Statistician*).

## Una nota de agradecimiento

Deseamos agradecer a Randy Craig, Salem State University; Mark Eakin, University of Texas-Arlington; Kathy Ernstberger, Indiana University-Southeast; Kimberley Killmer Hollister, Montclair State University; C.P. Kartha, University of Michigan, Flint; Robert Lemke, Lake Forest College; Ram Misra, Montclair State University; Prashant Palvia, University of North Carolina, Greensboro; Susan Parisien, Merrimack College; Brock Williams, Texas Tech University; Frederick Wiseman, Northeastern University; Reginald Worthley, University of Hawaii, Manoa; y Charles Zimmerman, Robert Morris College, por sus comentarios que contribuyeron a mejorar este libro.

Especialmente agradecemos a Debbie Clare, Mark Pfaltzgraff, Jeff Shelstad, Alana Bradley, Anne Graydon, Cynthia Regan, Nancy Welcher y Jane Avery de los equipos editoriales, de marketing y producción de Prentice Hall. Ha sido un privilegio trabajar con Tom Tucker en este proyecto así como en muchos proyectos previos. Como ahora Tom cambia de profesión, extrañaremos mucho su discernimiento, apoyo y dedicación. Gracias Tom y ¡buena suerte!

Queremos agradecer a nuestros lectores y revisores especialistas Annie Puciloski, Stonehill College y James Zimmer, Chatanooga State University, por su diligencia al revisar nuestro trabajo; a Robie Grant por su revisión de pruebas; a Julie Kennedy por su trabajo de edición, y a Sandra Krausman de GGS Book Services, Atlantic Highlands, por su trabajo en la producción de este texto.

Estamos muy agradecidos por el amor y apoyo de nuestras familias. Nuestros padres Reuben y Lee Levine, Marvin Krehbiel, Roberta Reed y Nat y Ethel Berenson, nos han bendecido con una vida de apoyo. Por último, quisiéramos agradecer a nuestras esposas e hijos por su paciencia, comprensión, amor y apoyo para hacer de este libro una realidad. Es a ellos a quienes dedicamos esta obra.

## Comentarios finales

Hemos llegado lejos para hacer este texto sólido desde el punto de vista pedagógico y libre de errores. Si tiene cualquier sugerencia o requiere de alguna aclaración sobre el material, o si encuentra algún error, por favor envíe un mensaje a [David\\_Levine@BARUCH.CUNY.EDU](mailto:David_Levine@BARUCH.CUNY.EDU) o [KREHBITC@MUOHIO.EDU](mailto:KREHBITC@MUOHIO.EDU). Incluya la frase BSFC—version 4 en el espacio de Asunto de su correo electrónico. Para mayor información acerca del uso de PHStat2, véase los apéndices F y G, y el archivo de lectura PHStat2 en el CD-ROM que acompaña este libro.

*David M. Levine*

*Timothy C. Krehbiel*

*Mark L. Berenson*

# CAPÍTULO 1

## Introducción y recolección de datos

---

**USO DE LA ESTADÍSTICA:** Good Tunes

- 1.1 CONCEPTOS BÁSICOS DE LA ESTADÍSTICA**
- 1.2 EL CRECIMIENTO DE LA ESTADÍSTICA Y LAS TECNOLOGÍAS DE LA INFORMACIÓN**
- 1.3 CÓMO ESTÁ ORGANIZADO ESTE TEXTO**
- 1.4 RECOLECCIÓN DE DATOS**  
Identificación de las fuentes de datos

- 1.5 TIPOS DE DATOS**
- A.1 INTRODUCCIÓN AL USO DE LOS PROGRAMAS ESTADÍSTICOS**
  - A1.1 Uso de Windows*
  - A1.2 Introducción a Excel*
  - A1.3 Introducción a Minitab*
  - A1.4 (Tema CD-ROM) Introducción a SPSS*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- Cómo se usa la estadística en los negocios
- Cuáles son las bases de datos que se utilizan en los negocios
- Los tipos de datos usados en negocios

# USO DE LA ESTADÍSTICA



## Good Tunes—Parte I

Good Tunes es un comerciante minorista en línea de sistemas de entretenimiento, que busca expandir su negocio abriendo varias tiendas. Para obtener el financiamiento que garanticé su expansión, Good Tunes requiere solicitar préstamos a los bancos locales. Los administradores de la firma acuerdan realizar una presentación con diapositivas que expliquen su negocio y muestren su situación real para convencer a los banqueros de prestarles el dinero que necesitan. A usted se le pide que ayude en el proceso de preparación de las diapositivas. ¿Qué hechos incluiría? ¿Cómo los presentaría?

**T**odos los días usted hace acopio de noticias e información que le sirven de guía en su vida. Así, escuchar el reporte meteorológico le ayudará a decidir qué ropa usar, y si vive en una ciudad grande tal vez deba escuchar el reporte vial que le indique la mejor ruta para llegar al trabajo o a la escuela.

Sus gustos personales, así como las cosas que no le agradan, también rigen algunas de sus decisiones. A pesar de las críticas negativas que escucha acerca de alguna película, tal vez decida verla sólo por la admiración que siente hacia cierto actor.

De igual forma, los administradores de empresas deben tomar decisiones todos los días. Aunque en ocasiones evidencian sus “impulsos viscerales” al tomar decisiones (lo que se conoce formalmente como toma de decisión no estructurada), la mayoría de las veces deciden a partir de hechos concretos. Como estudiante de negocios, usted no debe tomar decisiones no estructuradas, ya que éstas requieren de intuición y discernimiento que se desarrollan tras muchos años de experiencia. Sin embargo, puede aprender procedimientos y métodos que le ayudarán a tomar mejores decisiones basadas en hechos concretos. Cuando comience a familiarizarse con los procedimientos y métodos implicados en la recolección, la presentación y la elaboración de resúmenes de un conjunto de datos, o a obtener conclusiones acerca de tales datos, entonces habrá descubierto la estadística.

En el escenario de Good Tunes, usted debe partir de la suposición razonable de que los banqueros buscan tomar decisiones basadas en hechos concretos que usted presentará, y no en otros factores, tales como antojos, gustos o aversiones personales. Presentar información errónea o aun correcta pero de forma inadecuada, podría llevar a los banqueros a tomar decisiones administrativas equivocadas, lo cual pondría en riesgo el futuro de Good Tunes. Usted necesita saber de estadística para presentar los hechos necesarios, y para aprender estadística lo primero que debe conocer son sus conceptos básicos.

### 1.1 CONCEPTOS BÁSICOS DE LA ESTADÍSTICA

La **estadística** es la rama de las matemáticas que examina las formas de procesar y analizar datos. La estadística ofrece los procedimientos para recolectar y transformar los datos de manera que sean útiles a quienes toman decisiones en los negocios. Para comprender la estadística, primero necesita conocer la definición de una variable.

## VARIABLES

Las **variables** son las características de los objetos o de los individuos.

Ejemplos de variables son el género al que usted pertenece, su especialidad o campo de estudio, la cantidad de dinero que tiene en su cartera y el tiempo que le toma alistarse por la mañana para ir a la escuela. El aspecto esencial de la palabra *variable* es la idea de que las cosas difieren y las personas también. Tal vez la persona que está sentada a su lado sea hombre, quizás se esté especializando en algún campo de estudio totalmente diferente del suyo, de seguro tendrá una cantidad diferente de dinero en su cartera y sin duda ocupará una cantidad de tiempo diferente al suyo para alistarse e ir al trabajo cada mañana. Se debe distinguir entre una variable, como el género, y su *valor* para una observación individual (por ejemplo, “masculino”).

Todas las variables deben tener una **definición operacional**, es decir, un significado universalmente aceptado que sea claro para todos aquellos que estén relacionados con el análisis. La falta de definiciones operacionales genera confusión. Un ejemplo notable de confusión que ilustra la importancia de las definiciones operacionales es el de las elecciones presidenciales del año 2000 en Estados Unidos y las boletas electorales del estado de Florida que estuvieron en disputa (Jackie Calmes y Edward P. Foldessy, “In Election Review, Bush Wins with No Supreme Court Help”, *The Wall Street Journal*, 12 de noviembre de 2001, A1, A14). Con la ayuda del National Opinion Research Center de la Universidad de Chicago, se llevó a cabo una revisión de 175,010 boletas que fueron rechazadas porque no registraban votos presidenciales, o bien, porque contenían votos a favor de dos o más candidatos. Se utilizaron nueve estándares o definiciones operacionales para evaluar las boletas. Estos nueve estándares generaron diferentes resultados. Tres de los estándares (incluido uno propuesto por Al Gore) llevaron a George Bush a tener márgenes de victoria que iban de 225 a 493 votos. Seis de los estándares (incluido uno propuesto por George Bush) llevaron a Al Gore a tener márgenes de victoria que iban de 42 a 171 votos.

## POBLACIÓN

La **población** consiste en todos los miembros de un grupo acerca de los cuales se desea obtener una conclusión.

## MUESTRA

Una **muestra** es una parte de la población seleccionada para análisis.

## PARÁMETRO

Un **parámetro** es una medida numérica que describe una característica de la población.

## ESTADÍSTICO

Un **estadístico** es la medida numérica que describe alguna característica de la muestra.

Ahora que se han definido las variables, usted necesita comprender el significado de conceptos como población, muestra, parámetro y estadístico.

Ejemplos de poblaciones son todos los estudiantes de tiempo completo de una universidad, todos los votantes registrados en la ciudad de Nueva York, y la gente que fue de compras al centro comercial de alguna ciudad el pasado fin de semana. Las muestras podrían seleccionarse a partir de cada una de esas tres poblaciones. Los ejemplos incluyen 10 alumnos de tiempo completo seleccionados para participar en una investigación, 500 votantes registrados en Nueva York con los que se estableció comunicación vía telefónica para realizar una encuesta política, y 30 compradores del centro comercial encuestados sobre el grado de satisfacción del consumidor. En cada caso, la gente de la muestra representa una porción o subconjunto de la gente comprendida en la población.

La cantidad promedio que gastó la gente que fue de compras al centro comercial el fin de semana pasado es un parámetro. Se requiere información de todos los compradores para calcular este parámetro. La cantidad promedio que gastaron los 30 compradores que respondieron a la encuesta sobre el grado de satisfacción del consumidor es un estadístico. La información de estos 30 compradores se emplea para calcular el estadístico.

La estadística se divide en dos ramas, ambas aplicables a la administración de negocios. La **estadística descriptiva** se enfoca en la recolección, resumen y presentación de un conjunto de datos. La **estadística inferencial** utiliza datos de las muestras para obtener conclusiones acerca de cierta población.

La estadística descriptiva tiene sus raíces en la necesidad de las grandes organizaciones políticas y sociales de guardar registros. Por ejemplo, cada década desde 1790, en Estados Unidos se realiza un censo que recolecta y resume datos acerca de los ciudadanos. A través de los años, el U.S. Census Bureau ha sido uno de los grupos que ha mejorado los métodos de estadística descriptiva. Los fundamentos de la estadística inferencial se basan en las matemáticas de la teoría de la probabilidad. Los métodos inferenciales utilizan los datos de la muestra para calcular los estadísticos que proporcionan los estimados de las características de una población.

En la actualidad, los métodos estadísticos se aplican en diferentes áreas de los negocios. La contabilidad utiliza métodos estadísticos para seleccionar muestras con el propósito de auditar y para entender los costos de los conductores en la contabilidad de costos. Las finanzas usan métodos estadísticos para seleccionar entre alternativas de cuentas de inversión y para analizar las mediciones financieras a través del tiempo. La administración utiliza métodos estadísticos para mejorar la calidad de los productos manufacturados o los servicios que ofrece una organización. El marketing emplea métodos estadísticos para estimar el número de consumidores que prefieren un producto sobre otro y para saber por qué lo prefieren, así como para obtener conclusiones que permitan determinar la mejor estrategia de publicidad que incremente las ventas de un producto.

## 1.2 EL CRECIMIENTO DE LA ESTADÍSTICA Y LAS TECNOLOGÍAS DE LA INFORMACIÓN

Durante el siglo pasado, la estadística jugó un papel importante al promover el uso de la tecnología de la información y ésta, por su parte, contribuyó a difundir el uso de la estadística. Al iniciar el siglo XX, la expansión de los requerimientos de manejo de datos asociados con el censo federal llevó directamente al desarrollo de máquinas tabuladoras que fueron las antecesoras de los sistemas computacionales que utilizan los negocios en la actualidad. Especialistas como Pearson, Fisher, Gosset, Neyman, Wald y Tukey establecieron las técnicas de la estadística inferencial moderna, en respuesta a la necesidad de analizar grandes conjuntos de datos poblacionales que, ya para entonces, implicaban costos elevados, consumían mucho tiempo y eran difíciles de recopilar. El desarrollo de los primeros sistemas computacionales permitió a otros diseñar programas de cómputo que facilitaran el cálculo y procesamiento de datos impuestos por tales técnicas. A la vez, esos programas iniciales permitieron a los responsables de tomar decisiones hacer un mayor uso de los métodos estadísticos; por otra parte, los recientes avances en la tecnología de la información ayudaron a difundir el empleo de métodos estadísticos más complejos.

En la actualidad, cuando escuche acerca de comerciantes minoristas que invierten en “sistemas de administración de relaciones con el cliente”, o en un paquete productor de bienes dedicado al “data mining” para descubrir las preferencias de los consumidores, se dará cuenta de que las técnicas estadísticas son como los cimientos de esas aplicaciones de la tecnología de la información. Aun cuando tales aplicaciones pudieran requerir programación a la medida, durante muchos años los negocios han tenido acceso a **paquetes estadísticos**, tales como Minitab y SPSS, que son programas estandarizados para ayudar a los administradores a usar una amplia gama de técnicas estadísticas para automatizar el procesamiento y cálculo de datos que requieren esas técnicas. Mientras que en el pasado esos paquetes estaban disponibles sólo en centros computacionales corporativos, el aumento en el poder y la conectividad de las computadoras personales ha llevado estos paquetes al escritorio, donde se han reunido con herramientas familiares como los procesadores de palabras, las hojas de cálculo y los programas de la Web.

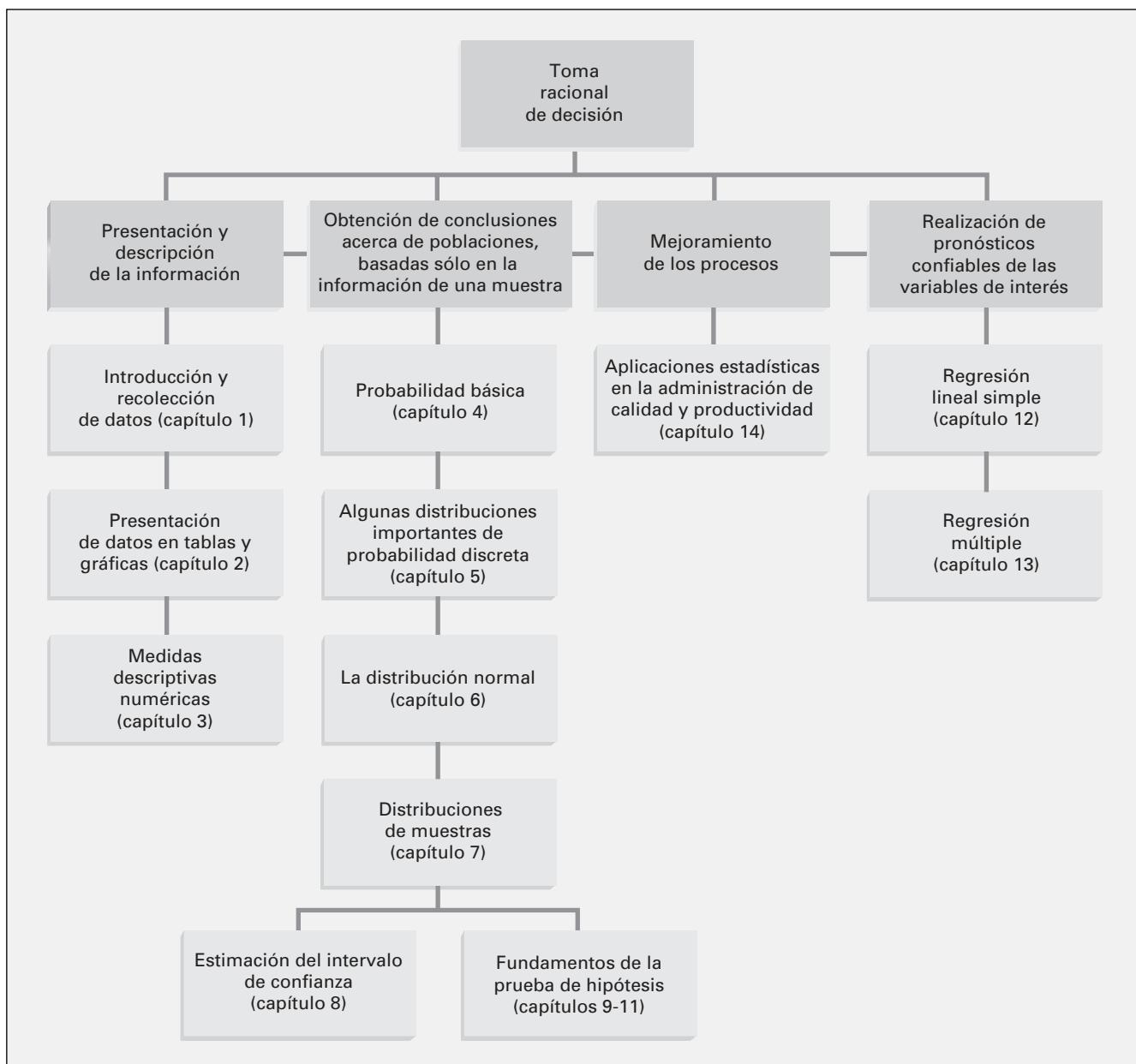
Los costos de arrendamiento y capacitación asociados a los paquetes estadísticos han llevado a muchos a considerar el uso de las funciones de gráficos y estadísticos del programa Excel, de Microsoft. Sin embargo, es necesario que sea cuidadoso de las preocupaciones que muchos especialistas en la materia tienen acerca de qué tan completos y precisos son los resultados estadísticos de Excel. Por desgracia, algunos investigadores han determinado que ciertas herramientas estadísticas del programa contienen fallas que pueden invalidar los resultados, en especial cuando los conjuntos de datos usados son muy grandes o tienen propiedades estadísticas poco usuales (vea la referencia 3 al final del capítulo). Está claro que cuando usted utilice Excel, deberá ser cuidadoso acerca de los datos y análisis que maneje. El hecho de que quizás las complicaciones sobreponen los beneficios de las atractivas características de Excel es todavía un enigma sin respuesta para los negocios.

## 1.3 CÓMO ESTÁ ORGANIZADO ESTE TEXTO

El objetivo principal de este texto es ayudarle a aprender y comprender cómo los métodos estadísticos resultan útiles en los procesos de toma de decisiones. Para los estudiantes de negocios, esta comprensión incluye los siguientes objetivos:

- Presentar y describir de forma apropiada los datos e información en los negocios.
- Llegar a conclusiones sobre grandes poblaciones únicamente a partir de información recolectada de las muestras.
- Hacer predicciones confiables acerca de las tendencias en los negocios.
- Mejorar los procesos de los negocios.

Este texto considera estos cuatro objetivos como principio organizador. La figura 1.1 muestra la forma en que cada capítulo se relaciona con estos objetivos. Usted explorará los métodos implicados en la recolección, presentación y descripción de información en lo que resta de este capítulo así



**FIGURA 1.1** Tabla estructural para este texto.

como en los capítulos 2 y 3. En los capítulos 4 a 7 aprenderá sobre los conceptos básicos de probabilidad y de las distribuciones binomial, normal y otras, que comprenderá mejor en los capítulos 8 a 11, donde se estudiará cómo obtener conclusiones acerca de grandes poblaciones a partir de la información de las muestras. En los capítulos 12 y 13 se enfocará en el análisis de la regresión que se utiliza para hacer pronósticos. En el capítulo 14 aprenderá métodos para mejorar los procesos en los negocios.

Como el aprendizaje contextual mejora la comprensión, cada capítulo inicia con un escenario de “Uso de la estadística” como el de “Good Tunes—Parte I” de la página 2, que identifica un problema de negocios en el que se aplica la estadística para convertir los datos en información útil para tomar una decisión racional. Las preguntas que surgen de los escenarios llevan a respuestas en forma de métodos estadísticos presentados en secciones subsiguientes del texto. Al reflexionar sobre estas preguntas, ganará en la apreciación de cómo los administradores de negocios usan la estadística en la actualidad para solucionar sus problemas y mejorar la calidad de sus productos y servicios.

Para el escenario de “Good Tunes—Parte I”, elegir lo que presentará es tan importante como elegir el método adecuado para la presentación y el resumen. En este caso, probablemente los banqueros pedirán conocer algunos datos, es decir, las “finanzas” del negocio. Pero, ¿qué otros datos recolectaría y presentaría usted para lograr la aprobación de los créditos? (Vea “Good Tunes—Parte II”.) Desde luego, después de haber presentado sus datos, usted esperaría que los banqueros hicieran las inferencias correctas. Es decir, esperaría que los banqueros fueran conocedores de los métodos estadísticos apropiados que conduzcan a la decisión de otorgar el crédito!

# USO DE LA ESTADÍSTICA

## Good Tunes—Parte II

Los propietarios de Good Tunes han decidido apoyar los datos financieros de su solicitud de préstamo, con los datos referentes a las percepciones de los clientes acerca del negocio Good Tunes. Para ayudar a evaluar estas percepciones, Good Tunes ha pedido a sus clientes responder y devolver una encuesta, que se incluye en cada pedido, acerca del grado de satisfacción con la empresa. La encuesta incluye las siguientes preguntas:

- ¿Cuántos días pasaron desde el momento en que ordenó su mercancía hasta que la recibió? \_\_\_\_\_
- ¿Cuánto dinero (en dólares estadounidenses) espera gastar en equipo electrónico y de estéreo en los próximos 12 meses? \_\_\_\_\_
- En general, ¿cómo calificaría el servicio que ofrece Good Tunes de acuerdo con su más reciente compra?
 

Mucho mejor de lo esperado	<input type="checkbox"/>	Peor de lo esperado	<input type="checkbox"/>
Mejor de lo esperado	<input type="checkbox"/>	Mucho peor de lo esperado	<input type="checkbox"/>
Como era de esperarse	<input type="checkbox"/>		
- ¿Cómo calificaría la calidad de los artículos que recientemente compró en Good Tunes?
 

Mucho mejor de lo esperado	<input type="checkbox"/>	Peor de lo esperado	<input type="checkbox"/>
Mejor de lo esperado	<input type="checkbox"/>	Mucho peor de lo esperado	<input type="checkbox"/>
Como era de esperarse	<input type="checkbox"/>		
- ¿Es probable que compre mercancía adicional a través de Good Tunes en los próximos 12 meses? Sí  No

A usted se le pide revisar esta encuesta. ¿Qué tipo de datos pretende recolectar? ¿Qué tipo de información es posible generar a partir de los datos obtenidos por la encuesta? ¿Cómo puede Good Tunes usar dicha información para mejorar la calidad del servicio y la mercancía? ¿Cómo podría Good Tunes utilizar la información para aumentar la probabilidad de obtener el préstamo? ¿Qué otras preguntas sugiere que se incluyan en la encuesta?

## 1.4 RECOLECCIÓN DE DATOS

Administrar un negocio de forma efectiva requiere la recolección de los datos apropiados. En muchas ocasiones, los datos son medidas que se obtienen de los elementos de una muestra, y las muestras se toman de la población, de tal forma que sean lo más representativas posible. La técnica más común para asegurar una representación adecuada es usar una muestra aleatoria. (Vea el capítulo 7 para una explicación más detallada de las técnicas de muestreo.)

Existen diversas circunstancias que requieren la recolección de datos:

- Un analista de investigación de mercados necesita evaluar la efectividad de una nueva campaña publicitaria en televisión.
- Un productor farmacéutico necesita determinar si un nuevo medicamento es más efectivo que los que actualmente se consumen.
- Un administrador de operaciones desea monitorear el proceso de producción para comprobar si la calidad de cierto producto satisface los estándares de la compañía.
- Un auditor desea revisar las transacciones financieras de una empresa para determinar si ésta cumple o no con principios contables aceptables.
- Un inversionista potencial desea determinar qué firmas industriales tienen mayor probabilidad de crecer de forma acelerada en un periodo de recuperación económica.

### Identificación de las fuentes de datos

Identificar las fuentes de datos apropiadas es un aspecto importante del análisis estadístico. Si los sesgos, ambigüedades u otro tipo de errores estropean los datos que son recolectados, ni siquiera los métodos estadísticos más complejos producirán una información precisa. Existen cuatro importantes fuentes de datos:

- Los que proporciona una organización o un individuo.
- Un experimento diseñado.
- Una encuesta.
- Un estudio observacional.

Las fuentes de datos se clasifican en **fuentes primarias** y **fuentes secundarias**. Cuando el recolector de datos es quien los usa para el análisis, la fuente es primaria. Cuando una organización o individuo han compilado los datos que utiliza otra organización o individuo, la fuente es secundaria.

Las organizaciones e individuos que recolectan y publican datos, generalmente los utilizan como fuente primaria y después permiten a otros usarlos como fuente secundaria. Por ejemplo, el gobierno federal de Estados Unidos recolecta y distribuye datos tanto para propósitos públicos como privados. El Bureau of Labor Statistics recolecta los datos que emplea y también distribuye cada mes el *Consumer Price Index*. El Census Bureau supervisa una gran variedad de encuestas actuales referentes a población, vivienda e industria, y lleva a cabo estudios especiales en temas como el crimen, los viajes y el cuidado de la salud.

Las empresas de investigación de mercado y las asociaciones de comercio también distribuyen datos referentes a industrias o mercados específicos. Los servicios de inversión como Mergent proporcionan datos financieros en una base de compañía a compañía. Servicios como los de A.C. Nielsen ofrecen a sus clientes datos que permiten la comparación de los productos del cliente con los de sus competidores. Todos los días los periódicos están repletos de información numérica referente a los precios de las acciones, las condiciones del clima y estadísticas deportivas.

Como expresamos antes, conducir un experimento es otra fuente importante de recolección de datos. Por ejemplo, para probar la efectividad de un detergente, un experimentador determina qué marcas son más eficientes para dejar limpia la ropa sucia, lavándola directamente, en lugar de preguntar a los clientes qué marca creen que sea más eficaz. Diseños experimentales adecuados generalmente son tema de estudio en textos más avanzados porque a menudo implican procedimientos estadísticos muy complejos. Sin embargo, consideraremos algunos conceptos fundamentales del diseño experimental en el capítulo 10.

Realizar una encuesta es la tercera fuente de datos importante. En ella se pregunta a la gente sobre sus creencias, actitudes, comportamientos y otras características. Las respuestas posteriormente se editan, codifican y tabulan para su análisis.

Dirigir un estudio observacional es la cuarta fuente de datos importante. En este estudio, el investigador observa el comportamiento de forma directa, generalmente en su ambiente natural. Los estudios observacionales tienen muchas formas en los negocios. Un ejemplo de ellos es el **grupo focal**, una herramienta de investigación de mercado que se utiliza para provocar respuestas no estructuradas ante preguntas abiertas. En un grupo focal un moderador dirige la discusión y los participantes responden a las preguntas. Otros tipos de estudios más estructurados implican dinámicas de grupo y construcción de consenso y el uso de numerosas herramientas del comportamiento organizacional como la lluvia de ideas, la técnica Delphi y el método del grupo nominal. Las técnicas de los estudios observacionales también se utilizan en situaciones en las que el esfuerzo de un grupo de trabajo o el mejoramiento en la calidad de los productos son los objetivos o metas de la administración.

## 1.5 TIPOS DE DATOS

Los **datos** son los valores observados de las variables, por ejemplo, las respuestas a una encuesta. Los especialistas desarrollan encuestas para lidiar con una gran variedad de variables diferentes. Como se ilustra en la figura 1.2, existen dos tipos de variables: categóricas y numéricas.

**FIGURA 1.2**

Tipos de variables.

Tipo de dato	Tipo de preguntas	Respuestas
Categórico	→ ¿Posee actualmente algunas acciones o bonos?	Sí <input type="checkbox"/> No <input type="checkbox"/>
Numérico	Discreto → ¿A cuántas revistas está suscrito actualmente?	_____ Número
	Continuo → ¿Cuánto mide?	_____ Pulgadas

Las **variables categóricas** producen respuestas categóricas, tales como sí o no. Un ejemplo es la respuesta a la pregunta “¿Posee actualmente algunas acciones o bonos?”, porque se limita a una respuesta de sí o no. Otro ejemplo es la respuesta a la pregunta de la encuesta de Good Tunes (presentada en la página 6) “¿Es probable que compre mercancía adicional a través de Good Tunes en los próximos 12 meses?” En ocasiones, las variables categóricas también generan más de dos posibles repuestas. Por ejemplo, “¿qué día de la semana es más probable que coma en un restaurante?”.

Las **variables numéricas** producen respuestas tales como la estatura en pulgadas. Otros ejemplos son: “¿cuánto dinero (en dólares estadounidenses) espera gastar en equipo electrónico y de estéreo en los próximos 12 meses?” (Tomado de la encuesta del grado de satisfacción del cliente de Good Tunes), o la respuesta a la pregunta “¿A cuántas revistas está suscrito actualmente?” Hay dos tipos de variables numéricas: discretas y continuas.

Las **variables discretas** producen respuestas numéricas que surgen de un proceso de conteo. “El número de revistas a las que está suscrito” es un ejemplo de una variable numérica discreta, porque la respuesta es una de un número finito de números enteros. Una persona se suscribe a cero, una, dos, o más revistas.

Las **variables continuas** producen respuestas numéricas que surgen de un proceso de medición. La estatura es un ejemplo de una variable numérica continua, porque la respuesta toma cualquier valor dentro de un continuo o intervalo, dependiendo de la precisión del instrumento que se utilice para medir. Por ejemplo, una persona podría medir 67 pulgadas,  $67\frac{1}{4}$  pulgadas,  $67\frac{7}{32}$  pulgadas o  $67\frac{58}{250}$  pulgadas, dependiendo de la precisión de los instrumentos disponibles.

No hay dos personas que midan exactamente lo mismo y cuanto más precisos sean los dispositivos de medición, mayor será la probabilidad de detectar las diferencias entre las estaturas. Sin embargo, la mayoría de los dispositivos de medición no son tan precisos como para detectar pequeñas diferencias. Por lo tanto, a menudo encontramos *observaciones enlazadas* a los datos experimentales o de encuesta, aun cuando la variable sea verdaderamente continua, y teóricamente todos los valores de una variable continua son diferentes.

## PROBLEMAS PARA LA SECCIÓN 1.5

### Aprendizaje básico

**1.1** Se venden tres diferentes bebidas en un restaurante de comida rápida: refrescos, té y café. Explique por qué el tipo de bebida es un ejemplo de variable categórica.

**1.2** En el restaurante de comida rápida se venden refrescos en tres tamaños: chico, mediano y grande. Explique por qué el tamaño de la bebida es una variable categórica.

**1.3** Suponga que mide el tiempo que le toma bajar de Internet un archivo de MP3.

a. Explique por qué el tiempo que le toma bajarlo es una variable numérica.

b. ¿Esta variable es discreta o continua?

### Aplicación de conceptos



**1.4** Para cada una de las siguientes variables, determine si es categórica o numérica. Si la variable es numérica, determine si es discreta o continua.

- a. Número de aparatos telefónicos por casa.
- b. Duración (en minutos) de la llamada de larga distancia más prolongada hecha cada mes.
- c. Si existe en la casa una línea telefónica conectada a un módem de computadora.
- d. Si hay un fax en la casa.



**1.5** La siguiente información se recolecta de estudiantes que salen de la librería del campus durante la primera semana de clases:

- a. Cantidad de tiempo que pasan de compras en la librería.
- b. Número de libros de texto comprados.
- c. Especialidad académica.
- d. Género.

Clasifique cada una de estas variables como categóricas o numéricas. Si la variable es numérica, determine si es discreta o continua.



**1.6** Para cada una de las siguientes variables, determine si la variable es categórica o numérica. Si la variable es numérica, indique si es discreta o continua.

- a. Nombre del proveedor de Internet.
- b. Cantidad de tiempo que navega en Internet por semana.
- c. Número de correos electrónicos recibidos en una semana.
- d. Cantidad de compras hechas en línea durante un mes.

**1.7** Para cada una de las siguientes variables, determine si es categórica o numérica. Si es numérica indique si la variable es discreta o continua.

- a. Cantidad de dinero gastado en ropa en el último mes.
- b. Tienda departamental preferida.
- c. Periodo de tiempo durante el cual es más probable que realice compras de ropa (entre semana durante el día, entre semana durante la noche, fin de semana).
- d. Número de pares de guantes de invierno que posee.

**1.8** Suponga que la siguiente información fue recolectada de la solicitud de préstamo hipotecario que hizo Robert Keeler al Metro County Savings and Loan Association:

- a. Pagos mensuales: \$1,427.
- b. Número de empleos en los últimos 10 años: 1.
- c. Ingreso salarial familiar anual: \$86,000.
- d. Estado civil: casado.

Clasifique cada una de las respuestas por el tipo de datos.

**1.9** Una de las variables que más a menudo se incluyen en las encuestas es el ingreso. A veces la pregunta es: “¿cuál es su ingreso (en miles de dólares)?” En otras encuestas, a quien responde se le pide que “Coloque una X en el círculo que corresponda a su nivel de ingreso” y se ofrecen varios rangos para elegir.

- a. En el primer formato, explique por qué el ingreso podría considerarse tanto discreto como continuo.
- b. ¿Cuál de los dos formatos preferiría usar al conducir una encuesta? ¿Por qué?

**1.10** Si dos estudiantes obtienen una puntuación de 90 en el mismo examen, ¿qué argumentos podría usar para mostrar que la variable subyacente (puntuación de la prueba) es continua?

**1.11** El director de investigación de mercados de una gran cadena de tiendas departamentales desea llevar a cabo una encuesta en el área metropolitana para determinar la cantidad de tiempo que las mujeres pasan comprando ropa en un mes.

- a. Describa tanto la población como la muestra de interés, e indique el tipo de datos que el director desearía recolectar.
- b. Desarrolle un primer borrador del cuestionario requerido, anotando series de tres preguntas categóricas y tres preguntas numéricas que usted considere apropiadas para esta encuesta.

## RESUMEN

En este capítulo ha estudiado la recolección y los diferentes tipos de datos que se usan en los negocios. En el escenario de “Uso de la estadística” se le pidió que revisara la encuesta de clientes diseñada por la compañía Good Tunes (vea la página 6). Las primeras dos preguntas generarán datos numéricos y las últimas tres producirán datos categóricos. Las respuestas a la primera pregunta (número de días) son discretas y las respues-

tas a la segunda pregunta (cantidad de dinero gastado) son continuas. Después de recolectar los datos, deben organizarse y prepararse para realizar varios análisis. En los siguientes dos capítulos se desarrollarán los temas de tablas, gráficas y una gran variedad de medidas numéricas que son útiles para analizar los datos.

## CONCEPTOS CLAVE

Datos	8	Fuentes primarias	7	Población	3
Definición operacional	3	Fuentes secundarias	7	Variables	3
Estadística	2	Grupo focal	8	Variables categóricas	8
Estadística descriptiva	4	Muestra	3	Variables continuas	8
Estadística inferencial	4	Paquetes estadísticos	4	Variables discretas	8
Estadístico	3	Parámetro	3	Variables numéricas	8

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**1.12** ¿Cuál es la diferencia entre una muestra y una población?

**1.13** ¿Cuál es la diferencia entre un estadístico y un parámetro?

**1.14** ¿Cuál es la diferencia entre la estadística inferencial y la descriptiva?

**1.15** ¿Cuál es la diferencia entre una variable categórica y una numérica?

**1.16** ¿Cuál es la diferencia entre una variable discreta y una continua?

**1.17** ¿Qué es una definición operacional y por qué es importante?

### Aplicación de conceptos

**1.18** La Data and Story Library ([lib.stat.cmu.edu/DASL](http://lib.stat.cmu.edu/DASL)) es una librería de archivos de datos e historias en línea, que ilustra el uso de los métodos estadísticos básicos. Las historias están clasificadas por método y por tema. Visite este sitio y dé clic en **List all topics**. Seleccione una historia y elabore un resumen de cómo se ha utilizado la estadística en ella.

**1.19** Visite el sitio oficial de Microsoft Excel ([www.microsoft.com/office/excel](http://www.microsoft.com/office/excel)). Explique cómo cree que sea útil el programa Excel en el campo de la estadística.

**1.20** Visite el sitio oficial de Minitab Web ([www.minitab.com](http://www.minitab.com)). Explique cómo cree que Minitab resulte útil en el campo de la estadística.

**1.21** Visite el sitio oficial de SPSS ([www.spss.com](http://www.spss.com)). Explique cómo cree que sea útil el SPSS en el campo de la estadística.

**1.22** La organización Gallup pone a disposición los resultados de las elecciones en su sitio Web ([www.gallup.com](http://www.gallup.com)). Visítelo y dé clic en algún artículo que le interese de la sección “Top Stories”.

- Dé un ejemplo de una variable categórica que se encuentre en el artículo.
- Dé un ejemplo de una variable numérica encontrada en el artículo.
- La variable que seleccionó en b), ¿es discreta o continua?

**1.23** El sitio del U.S. Census Bureau ([www.census.gov](http://www.census.gov)) contiene la información de las encuestas sobre población, negocios, geografía y otros temas. Visite el sitio, dé clic en **Housing**, en la sección “People”. Después dé clic en **American Housing Survey**.

- Describa brevemente el American Housing Survey.
- Dé un ejemplo de una variable categórica que haya encontrado en la encuesta.
- Dé un ejemplo de una variable numérica que haya encontrado en la encuesta.
- La variable que seleccionó en c), ¿es discreta o continua?

**1.24** En el sitio Web del U.S. Census Bureau ([www.census.gov](http://www.census.gov)), dé clic en **Survey of Business Owners** en la sección “Business” y lea la descripción The Survey of Business Owners and Self-Employed Persons (SBO). Dé clic en **SBO-1**, en la sección “Forms and Instructions” para visualizar la forma de encuesta usada.

- Dé un ejemplo de una variable aleatoria categórica encontrada en esta encuesta.
- Dé un ejemplo de una variable aleatoria numérica encontrada en esta encuesta.
- La variable que seleccionó en b), ¿es discreta o continua?

**1.25** En un reporte basado en un estadístico del Departamento del Transporte de Estados Unidos U.S., la empresa JetBlue fue la número 1 en calidad entre todas las líneas aéreas de ese país en 2003. JetBlue obtuvo el segundo mejor tiempo de desempeño, al llegar a tiempo el 86% de las veces. También los clientes de JetBlue llenaron menos formatos de quejas que los de todas las demás líneas aéreas, salvo una (“JetBlue calificada como la línea aérea número 1, dice el reporte”, [USAToday.com](http://USAToday.com), 5 de abril, 2004).

- ¿Cuál o cuáles de los cuatro tipos de fuentes de datos listados en la sección 1.4 en la página 7 cree que se utilizaron en este estudio?
- Nombre una variable categórica usada en este artículo.
- Nombre una variable numérica usada en este artículo.

**1.26** De acuerdo con una encuesta de Goldman Sachs, sólo cerca del 4% de los hogares estadounidenses utilizan los servicios bancarios *on line*. Una encuesta realizada por Cyber Dialogue investigó las razones por las que la gente abandona el banco *on line* después de intentarlo. A continuación se ofrece un listado parcial de los resultados obtenidos por la encuesta

Cyber Dialogue (“USA Snapshots”, *USA Today*, 21 de febrero de 2000, A1).

#### ¿Por qué abandonó el banco *on line*?

Demasiado complicado o se lleva demasiado tiempo	27%
No le satisface el servicio al cliente	25%
No lo necesita/no está interesado	20%
Tiene preocupaciones acerca de la seguridad o de un fraude	11%
Demasiado costoso	11%
Tiene dudas sobre la privacidad	5%

**a.** Describa la población de la encuesta Goldman Sachs.  
**b.** Describa la población de la encuesta Cyber Dialogue.  
**c.** La respuesta a la pregunta “¿Por qué abandonó el banco *on line*?", es ¿categórica o numérica?

**d.** El 27% de quienes respondieron indicaron que el banco en línea era demasiado complicado o tomaba demasiado tiempo. ¿Es esto un parámetro o un estadístico?

**1.27** Un fabricante de comida para gatos planeaba realizar una encuesta en los hogares estadounidenses para determinar los hábitos de compra de quienes poseen gatos. Entre las preguntas que se incluirían estaban las relacionadas con la siguiente información:

1. Dónde se compra la comida de gato por lo general.
2. La comida de gato comprada es seca o húmeda.
3. El número de gatos que viven en el hogar.
4. Si el gato tiene o no pedigree.

- a. Describa la población.
- b. Para cada uno de los reactivos listados, indique si la variable es categórica o numérica. Si es numérica, ¿es discreta o continua?
- c. Plantee cinco preguntas categóricas para la encuesta.
- d. Plantee cinco preguntas numéricas para la encuesta.

## INTRODUCCIÓN A LOS CASOS WEB

### APRENDIENDO DE LOS CASOS WEB EN ESTE TEXTO

*La gente usa técnicas estadísticas para ayudar a comunicar y a presentar información importante a otros tanto dentro como fuera de sus negocios. Sin embargo, cada día, la gente hace mal uso de esas técnicas:*

- Un administrador de ventas que trabaja con un programa de gráficas “fácil de usar”, elige una gráfica inapropiada que oscurece las relaciones de los datos.
- El editor de un reporte anual presenta la gráfica de los ingresos con un eje *Y* reducido que crea la falsa impresión de que ha habido un gran incremento en los ingresos.
- Un analista genera estadísticos sin significado acerca de un conjunto de datos categóricos al usar un tipo de análisis diseñado para datos numéricos.

Aunque la mayoría de los casos en que se hace mal uso de la estadística no son intencionales, usted debe ser capaz de identificar todos esos usos incorrectos para ser un administrador bien informado. El objetivo principal de los Casos Web a lo largo del texto, es ayudarle a desarrollar este tipo de habilidad.

Se le pedirá que visite los sitios Web que están relacionados con las compañías y los temas tratados en el escenario de “Uso de la estadística” con el que inicia cada capítulo, o una página Web que apoye la continuación de la historia del *Springville Herald*, un diario de una pequeña ciudad. Revise también documentos internos así como demandas públicas, para identificar y corregir los usos incorrectos de la estadística. De forma diferente a como lo haría un libro de texto tradicional, pero apegado a las situaciones del mundo

real, no toda la información que encontrará será relevante para su tarea; en ocasiones descubrirá información conflictiva que tendrá que aclarar antes de continuar con los casos.

Para ayudarle con su aprendizaje, los Casos Web de cada capítulo inician con el objetivo de aprendizaje y la sinopsis del escenario que está bajo estudio. Se le dirigirá hacia un sitio o página Web específica y se le hará una serie de preguntas que guiarán su exploración. Si lo prefiere, también podrá explorar páginas Web para casos enlazados a la página del Springville Chamber of Commerce ([www.prenhall.com/Springville/SpringvilleCC.htm](http://www.prenhall.com/Springville/SpringvilleCC.htm)).

Para complementar a los Casos Web, en casi todos los capítulos encontrará un ejercicio tradicional de estudio de caso en el que se le pedirá que aplique sus conocimientos sobre la estadística en un problema que enfrente el *Springville Herald*.

Para ilustrar el uso de los Casos Web, visite el sitio de Good Tunes ([www.prenhall.com/Springville/Good\\_Tunes.htm](http://www.prenhall.com/Springville/Good_Tunes.htm)), el minorista en línea mencionado en los escenarios de “Uso de la estadística” en este capítulo. Recuerde que la empresa privada Good Tunes busca financiamiento para expandir su negocio abriendo otras tiendas. Como el interés del administrador es hacer que los banqueros vean en Good Tunes un negocio floreciente, no sorprende que utilice el argumento de que “éste ha sido nuestro mejor año en ventas”, como título de la página “Buenos tiempos para Good Tunes”.

Esta frase también es un hipervínculo, así es que dé un clic en “éste ha sido nuestro mejor año en ventas” para abrir la página que apoya esa afirmación. ¿Qué tan bien la apoya? ¿Con una tabla de números? ¿Con una gráfica? ¿Con declaraciones atribuidas a fuentes conocedoras? Good Tunes ha usado una gráfica para presentar los datos de sus ventas por

categoría “hace dos años” y “los últimos 12 meses”. ¿Hay problemas en las elecciones hechas en esta página Web? *Absolutamente!*

Primero, observe que no hay escalas para los símbolos usados, por lo que es imposible saber cuál es el volumen de ventas real. De hecho, como aprenderá en la sección 2.6, las gráficas que incorporan símbolos de esta forma se consideran como *gráficas basura*, que jamás utilizarían los especialistas.

Además de este importante punto, surge otra pregunta acerca de si los datos de venta representan el número de unidades vendidas o algo más. El uso de símbolos da la impresión de que se están presentando datos de unidades vendidas. Si es así, ¿apoyan estos datos el argumento, o tal vez otros datos, como el volumen en dólares, serían un mejor indicador de las ventas de Good Tunes?

Además están las curiosas etiquetas de las gráficas. “Los últimos 12 meses” es ambiguo —podría incluir meses del año en curso así como meses del año anterior y, por lo tanto, no ser equivalente a un periodo de “hace dos años”. Puesto que el negocio se estableció en 1997 y la categoría es “éste ha sido nuestro mejor año en ventas”, ¿por qué la administración no incluyó las cifras de ventas de *cada año*?

¿Está la administración de Good Tunes ocultando algo, o simplemente no son conscientes del uso adecuado de la estadística? De cualquier modo, no han logrado comunicar de forma adecuada un aspecto vital de su “historia”.

En los siguientes Casos Web se le pedirá realizar este tipo de análisis, usando las preguntas abiertas de este texto como guía. No todos los casos son tan claros y sinceros como esta muestra, y algunos casos incluyen aplicaciones estadísticas perfectamente apropiadas.

## REFERENCIAS

1. Kendall, M.G. y R.L. Plackett, eds., *Studies in the History of Statistics and Probability*, vol. 2 (Londres: Charles W. Griffin, 1977).
2. Kirk, R.E., ed., *Statistical Issues: A Reader for the Behavioral Sciences* (Monterey, CA: Brooks/Cole, 1972).
3. McCullough, B.D. y B. Wilson, “On the accuracy of statistical procedures in Microsoft Excel 97”, *Computational Statistics and Data Analysis*, 31 (1999), 27-37.
4. *Microsoft Excel 2003* (Redmond, WA: Microsoft Corporation, 2002).
5. *Minitab Release 14* (State College, PA: Minitab, Inc., 2004).
6. Pearson, E. S., ed., *The History of Statistics in the Seventeenth and Eighteenth Centuries* (Nueva York: Macmillan, 1978).
7. Pearson, E.S. y M.G. Kendall, eds., *Studies in the History of Statistics and Probability* (Darien, CT: Hafner, 1970).
8. *SPSS® Base 12.0 Brief Guide* (Upper Saddle River, NJ: Prentice Hall, 2003).

# Apéndice 1 Introducción del uso de los programas estadísticos

## ALGO SOBRE ESTE APÉNDICE

**Sección A1.1** Lea esta sección si no se encuentra familiarizado con las operaciones básicas de Windows que se requieren para usar de forma eficiente Excel, Minitab o SPSS.

**Sección A1.2** Lea esta sección sólo si planea usar Excel con este texto.

**Sección A1.3** Lea esta sección sólo si está usando Minitab con este texto.

**Sección A1.4** Lea esta sección del CD-ROM sólo si está usando el SPSS con este texto.

Nota: A lo largo de este apéndice y de los demás que se incluyen en este texto, el símbolo → se usa para representar una secuencia en el menú de selecciones. Por ejemplo, la instrucción “seleccionar **Archivo** → **Abrir**” significa que primero de-

be seleccionar el **Archivo** en el menú de opciones y posteriormente la opción **Abrir** del submenú que aparece.

### A1.1 USO DE WINDOWS

#### Uso del ratón

En Windows, con frecuencia usted usa el ratón u otro dispositivo señalador para seleccionar y activar opciones en la pantalla al presionar un botón. Por convención, Windows da por hecho que los dispositivos señaladores poseen dos botones, uno designado como botón primario y otro como botón secundario. Es posible mover el ratón, presionar y soltar sus botones de las siguientes maneras:

**Hacer clic o seleccionar:** Mover el ratón sobre un objeto y presionar el botón primario.

**Arrastrar:** Mover el ratón sobre un objeto. Mientras se presiona el botón primario, mover la flecha del ratón hacia otro lugar en la pantalla y soltar el botón. Al arrastrar se mueven los objetos hacia otra parte o también se pueden seleccionar varios puntos.

**Hacer doble clic:** Mover el ratón sobre un objeto y dar dos clics en el botón primario en una sucesión rápida.

**Hacer clic en el botón derecho:** Mover el ratón sobre un objeto y dar un clic en el botón secundario.

Por default, Windows define el botón izquierdo del ratón como botón primario y el derecho como el botón secundario (esto da lugar a la frase “clic-derecho”), pero es posible invertir estas definiciones a través del ícono Panel de control.

## Programas de apertura

Usted puede elegir una de dos formas para abrir directamente programas como Excel, Minitab o SPSS. Éstas formas son:

- **Clic en el ícono del programa:** Dar doble clic en el ícono del escritorio que representa el programa que utilizará (algunas versiones de Windows sólo requieren un clic).
- **Opción Inicio Menú:** Presione la tecla **con el ícono de Windows** (o dé clic en el botón Inicio en la pantalla) y elija la opción **Programa o Todos los programas**. De la lista que aparece, seleccione el programa que utilizará. Si éste aparece en un submenú, primero debe seleccionar el submenú y después el programa.

Los usuarios expertos de Windows conocen otras formas de abrir un programa. Usted puede, claro está, usar también alguna de esas formas, si así lo prefiere.

## Sistema operativo Windows

Windows debe su nombre al hecho de que cada vez que se abre un programa, también se abre una pantalla rectangular enmarcada o “ventana”. En ella se interactúa con el programa, lo que permite abrir ventanas adicionales en las que es posible entrar y realizar diversas actividades.

La ventana de apertura de la mayoría de los programas contiene los siguientes elementos comunes:

Una **barra de título** en la parte superior, que identifica el programa y cualquier archivo que esté en uso.

Los **botones para modificar el tamaño** en el lado derecho del área de la barra de título, que afectan el tamaño en que se exhibe la pantalla.

El **botón de cerrar el programa** en el margen superior derecho de la barra de título, que permite dar por terminado de forma rápida el uso del programa.

La **barra del menú** es una lista horizontal de palabras debajo de la barra de título que contiene el comando de opciones en uno o más menús.

Una o más **barras de herramientas** con botones que son comandos de acceso rápido.

La figura A1.1 muestra estos elementos comunes para abrir ventanas en Excel 2003, Minitab Release 14, y el SPSS Student Version 12. Note que las ventanas para estos programas adicionalmente tienen un área destinada a una hoja de trabajo, compuesta por filas y columnas que sirven para introducir datos.

## Ventanas de diálogo

Muchas entradas u opciones disponibles para utilizar el programa muestran ventanas adicionales conocidas como ventanas de diálogo. Las de Excel 2003 permiten abrir o imprimir un archivo (vea la figura A1.2), y contienen los siguientes elementos en común:

**Botón de ayuda con el signo de interrogación:** Dar clic en este botón le permite seleccionar un elemento de la ventana de diálogo y desplegar un mensaje de ayuda en cuanto a su uso.

**Cuadro de lista hacia abajo:** Al hacer clic en el botón se despliega hacia abajo una lista de opciones, que aparece en la esquina derecha de la ventana.

**Cuadro de lista:** Despliega una lista de opciones. En ocasiones incluye una **barra de desplazamiento** o **guía** si la lista es más grande de lo que el tamaño del cuadro podría mostrar.

**Cuadro de texto:** Ofrece un espacio para escribir una entrada. En ocasiones estos cuadros se combinan con una lista hacia abajo o con **botones** (que se observan en la página de la ventana en la figura A1.2), con formas alternativas de especificar una entrada.

**Cuadro de exploración:** Ofrece una serie de opciones para elegir desde cero, una o más de una opción (compare con los botones de opción).

**Iconos:** Permiten un acceso directo a otras aplicaciones en Windows en las que se pueden almacenar los archivos.

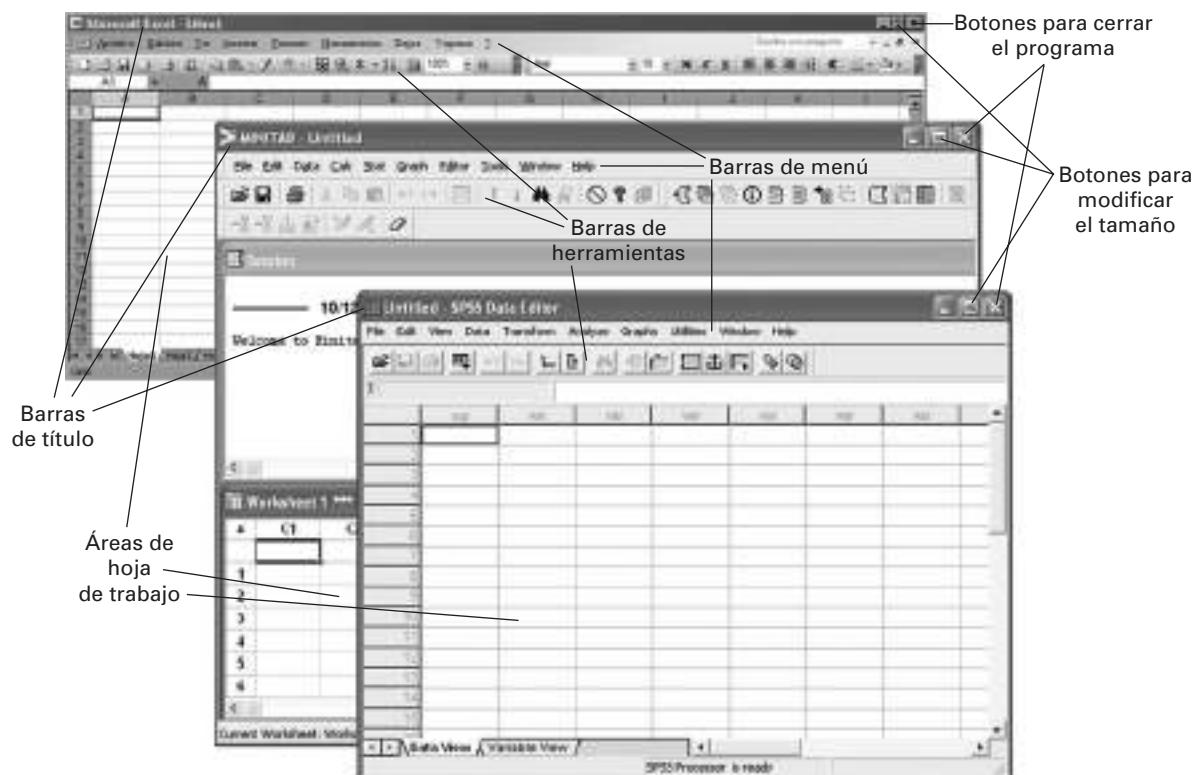
**Botones de opción:** Muestra un conjunto de elecciones mutuamente excluyentes en las cuales sólo una opción puede elegirse a la vez.

**Botones de comando:** Permite al programa realizar alguna acción que generalmente cierra la ventana de diálogo actual y muestra una adicional. El **botón Aceptar** hace que el programa actúe usando los valores y escenarios en uso de la ventana de diálogo. El **botón Cancelar** cierra el cuadro de diálogo y cancela la operación asociada con él.

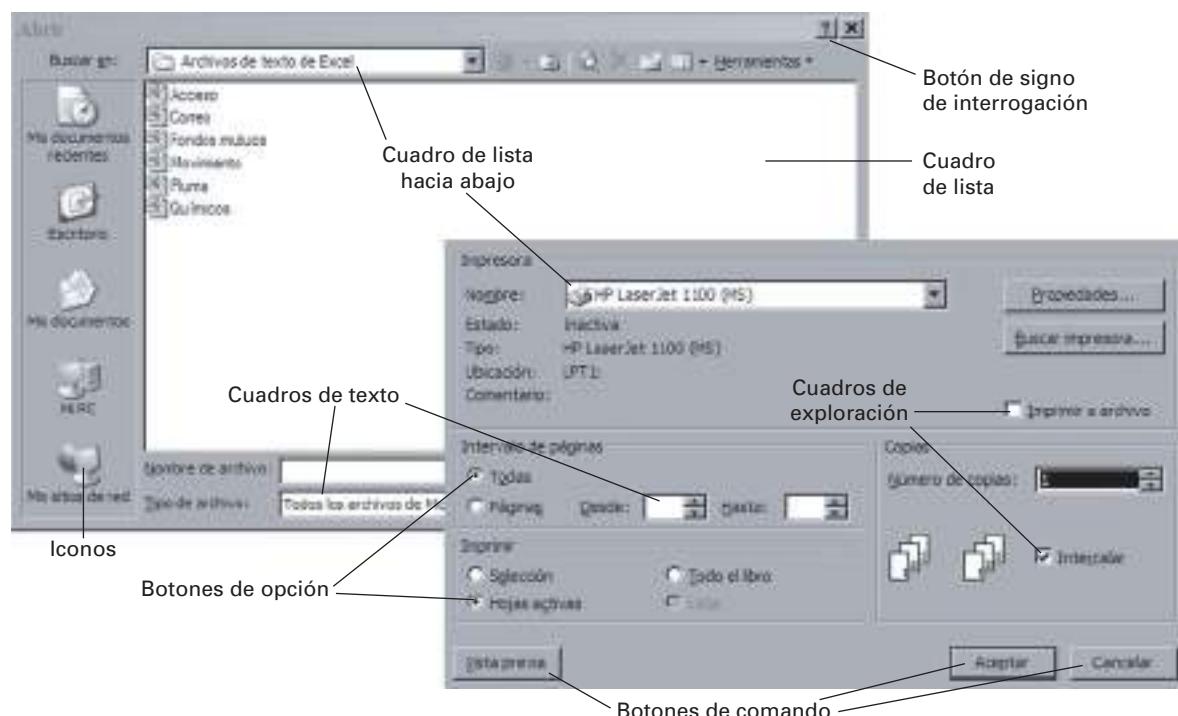
## A1.2 INTRODUCCIÓN A EXCEL

### Visión general de Excel

Excel es un programa de hoja de trabajo electrónica de Microsoft Office. Aunque no es un programa estadístico especializado, Excel contiene las funciones estadísticas básicas e incluye el paquete de herramientas para análisis de datos útiles para ejer-



**FIGURA A1.1** Ventanas de los programas Microsoft Excel, Minitab y SPSS.



**FIGURA A1.2** Elementos comunes a las ventanas de diálogo.

cutar métodos estadísticos avanzados. Puede instalar PHStat2 add-in de Prentice Hall (incluido en el CD-ROM que acompaña a este texto), el cual le permitirá extender y reforzar el paquete de herramientas para análisis de datos que contiene Microsoft. (No es necesario usar el PHStat2 para emplear Excel con este texto, aunque al hacerlo simplificará el uso de Excel para realizar los análisis estadísticos.)

Con Excel es posible crear o abrir y guardar archivos que se llaman **hojas de trabajo**. Éstos son conjuntos de hojas de trabajo y objetos relacionados, tales como gráficas que contienen datos originales, así como cálculos y resultados asociados con uno o más análisis. Por la amplia difusión de Excel, resulta conveniente usarlo, pero algunos especialistas en estadística han manifestado su preocupación en relación con el hecho de que arroja resultados que no son totalmente confiables y precisos para algunos procedimientos estadísticos. Aunque Microsoft ha mejorado muchas de sus funciones estadísticas a partir de Excel 2003, es necesario ser precavidos en su uso para análisis de datos, además de los que presenta este texto. (Si piensa instalar el PHStat2, asegúrese de leer primero el apéndice F y el archivo "Léame" de PHStat2 que aparecen en el CD-ROM incluido.)

## Uso de las hojas de trabajo de Excel

En Excel se introducen los datos en las hojas de trabajo que están organizadas en columnas tituladas con letras y filas numeradas. Generalmente se escriben los datos para cada variable en columnas separadas, en la celda de la fila uno para la etiqueta de una variable y cada fila subsiguiente para una observación singular. Se debe seguir la buena costumbre de introducir sólo un conjunto de datos por cada hoja de trabajo.

Para referirse a una entrada específica, o celda, se usa una notación *Sheetname!ColumnRow*. Por ejemplo, la notación Data!A2 se refiere a la celda de la columna A y la fila 2 en la hoja de trabajo. Para referirse a un grupo específico o **rango** de celdas, se usa una notación *Sheetname!Uppercell:Lowercell*. Por ejemplo, la notación Data!A2:B11 se refiere a las 20 celdas que están en las filas 2 a 11 en las columnas A y B de la hoja de trabajo.

Cada hoja de trabajo de Excel tiene su propio nombre. Excel las nombra de forma automática como **Hoja1**, **Hoja2**, etcétera. El usuario debería renombrar las hojas de cálculo con nombres más descriptivos. Esto se hace dando doble clic en las pestañas de las hojas que aparecen en la parte baja, se escribe el nuevo nombre y se presiona la tecla Enter para aceptar.

## Uso de fórmulas en las hojas de trabajo de Excel

Las fórmulas son entradas en las celdas de las hojas de trabajo, que ejecutan cálculos o algunas otras tareas. Se ingresan las fórmulas escribiendo el signo igual (=), seguido por una combinación de operaciones matemáticas o de procesamiento de datos.

Para las fórmulas simples se utilizan los símbolos +, -, \*, /, y ^ para las operaciones de suma, resta, multiplicación, división y exponente (un número elevado a una potencia), respectivamente. Por ejemplo, la fórmula =Data!B2 + Data!B3 + Data!B4 +

Data!B5 suma los contenidos de las celdas B2, B3, B4 y B5 de la hoja de trabajo y muestra la suma como el valor en la celda correspondiente. También se usan las *funciones* de Excel para simplificar las fórmulas. Por ejemplo, la fórmula =SUM(Data!B2:B5), que usa la función Excel SUM(), es el equivalente más corto de la fórmula de la oración anterior. También se usan las referencias de las celdas o rangos de celdas que no contienen la parte *Sheetname!*, tales como B2 o B2:B5. Estas referencias *siempre* se relacionan con la hoja de trabajo en la que se ingresó la fórmula.

Las fórmulas permiten crear soluciones generalizadas y darle a Excel su capacidad distintiva de volver a calcular de forma automática los resultados al cambiar los valores. Por lo general, al usar una hoja de trabajo sólo se ven los resultados de las fórmulas ingresadas, pero no las fórmulas en sí. Sin embargo, para su referencia en este texto se han incluido muchas ilustraciones de las hojas de trabajo de Excel que muestran las fórmulas subyacentes junto con los resultados que producen. Al usar Excel, es posible seleccionar **Herramientas** → **Opciones** y en la etiqueta **Vista** de la ventana de diálogo **Opciones** que aparece enseguida seleccionar el cuadro de exploración de **Fórmulas** y dar clic en el botón **Aceptar** para ver en la pantalla las fórmulas y no los resultados. Para restaurar la vista original, hay que deseleccionar el cuadro de exploración de **Fórmulas**.

## Uso del Asistente para gráficos de Excel

El Asistente para Gráficos de Excel permite generar una gran variedad de éstos. Éste es uno de los varios **Asistentes** de Microsoft Office, a través del cual un conjunto de ventanas de diálogo interrelacionadas nos guían paso a paso, conforme se crea un gráfico. Para usarlo, primero seleccione **Insertar** → **Gráfico**. Despues seleccione la información de las propiedades del gráfico conforme avanza a través de las ventanas de diálogo dando clic al botón **Siguiente**. Al dar clic en el botón **Finalizar** en la última ventana de diálogo, se termina con el Asistente y se crea el gráfico. Es posible cancelar la operación del Asistente en cualquier punto dando clic en el botón **Cancelar** o moviéndose hacia una ventana de diálogo previa dando clic en el botón **Atrás**.

Los Asistentes gráficos de las diferentes versiones de Excel difieren ligeramente. Para Excel 2003 el Asistente de gráficos de cuatro pasos (vea la figura A1.3) requiere lo siguiente:

**Paso 1:** Elija el tipo de gráfico.

**Paso 2:** Ingrese la ubicación de los datos, para los valores que serán graficados y la fuente de datos para los títulos informativos de la gráfica (si es que los hay).

**Paso 3:** Especifique el formato y las opciones de las etiquetas para el gráfico. (Vea más comentarios a la derecha.)

**Paso 4:** Elija la ubicación del gráfico en la hoja de trabajo. Usted creará un gráfico con una mejor escala si elige la opción "en una hoja nueva" en lugar de la opción "como objeto (en la hoja de trabajo)".

Es factible cambiar esta configuración después de que se haya producido el gráfico dando un clic con el botón derecho del ratón.

tón en el gráfico y haciendo la selección adecuada en el menú que aparece. Por ejemplo, para reconsiderar la colocación asociada con el paso 3 de la ventana de diálogo, debe seleccionar de ese menú **Opciones de gráfico**.

La configuración automática de la ventana de diálogo del paso 3 crea gráficos diseñados de forma imperfecta. Cuando usted usa el Asistente de gráficos para generar aquellos que requieren los problemas y ejemplos de este texto, debe seleccionar las pestañas que muestran abajo (vea la figura A1.4) y continuar las instrucciones (si una pestaña no aparece en la ventana de diálogo del paso 3 para un tipo específico de gráfico, ignore la instrucción de la pestaña):

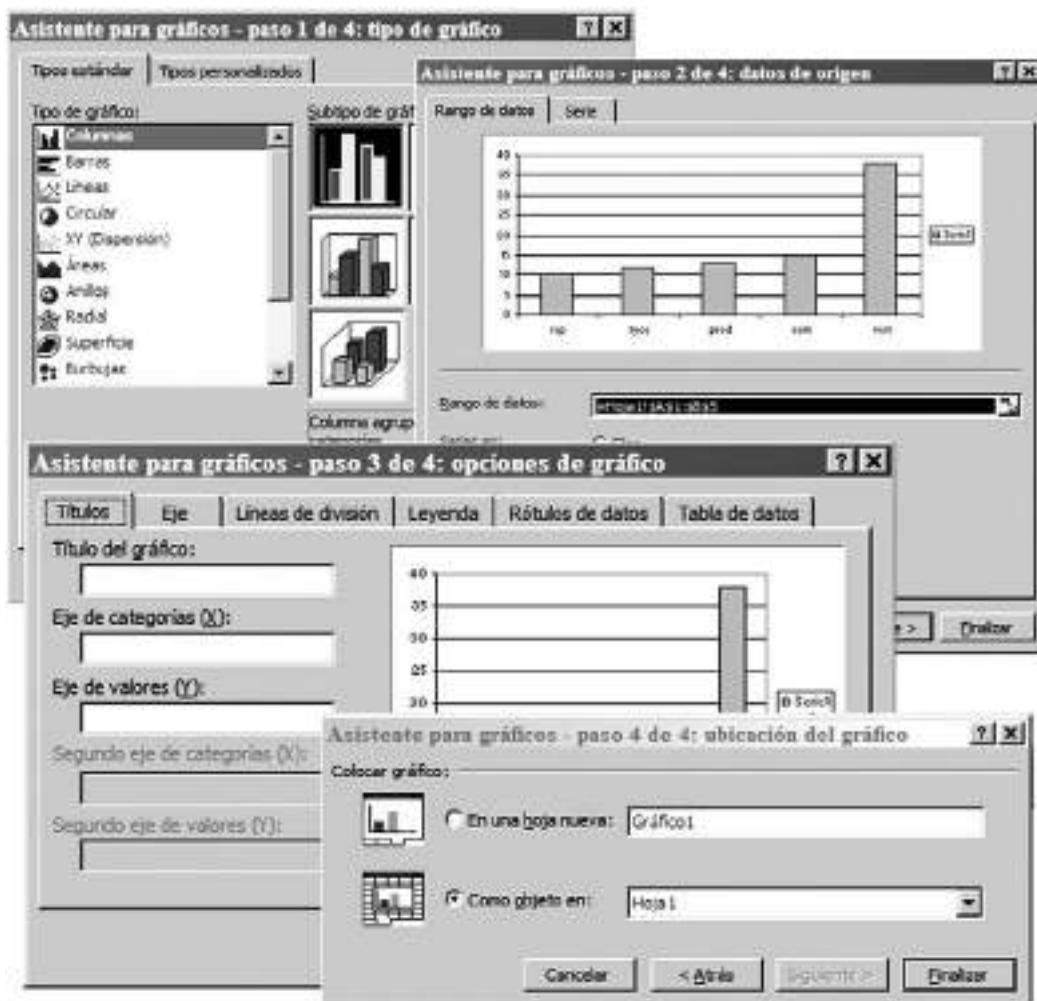
- Seleccione la pestaña **Títulos** e ingrese etiquetas para título y ejes, si es apropiado.
- Seleccione la pestaña **Ejes** y después en la ventana de exploración tanto el **eje (X)** como el **eje (Y)**. También seleccione el botón de la opción **Automático** bajo el cuadro de exploración del eje (X).
- Elija la pestaña **Líneas de división** y deseccione todas las opciones debajo del encabezado del eje (X) y bajo el encabezado del eje (Y).

- Seleccione la pestaña **Leyenda** y deseccione el cuadro de exploración de **Mostrar la leyenda**.
- Seleccione la pestaña **Rótulos de datos** y en esa pestaña seleccione el botón de la opción **Ninguno**.

### Abrir y guardar hojas de trabajo

Se abren las hojas para usar datos y resultados que usted u otras personas hayan creado antes. Para abrir una hoja de Excel, primero seleccione **Archivo → Abrir**. En la ventana de diálogo de Abrir que aparece, debe dar clic en el botón de **Aceptar**. Si no encuentra su archivo, tal vez necesitará hacer una o dos cosas más como:

- Utilizar las barras de desplazamiento, si se muestran, para deslizarse por la lista completa de archivos.
- Seleccionar el archivo correcto de la lista **Ver en** de la parte superior de la ventana de diálogo.
- Cambiar el valor de los **Tipos de archivos** de la lista hacia abajo en la parte inferior de la ventana de diálogo. Seleccionar la opción **Archivos de texto** de la lista para ver



**FIGURA A1.3** Ventanas de diálogo del Asistente para gráficas de Excel.



**FIGURA A1.4** Asistente de gráficos. Paso 3, pestañas de la ventana de diálogo.



**FIGURA A1.5** Ventana de diálogo para abrir de Excel.

cualquiera de los archivos de texto; para visualizar la lista de cada archivo del folder, seleccione **Todos los archivos**.

Para guardar una hoja de trabajo, seleccione **Archivo** → **Guardar como**, que le mostrará una ventana de diálogo similar a la de Abrir. Ingrese (o edite) el nombre del archivo en el cuadro **Nombre del archivo** y dé clic en el botón **Aceptar**. Si es aplicable, podría hacer algo de lo siguiente:

- Cambiar a otro folder seleccionando el de la lista hacia abajo de **Guardar en**.
- Cambiar el valor de **Guardar como tipo** en lugar de la opción que aparece por default, **Hojas de trabajo de Excel**. Las funciones “**Texto (etiqueta delimitada)**” o “**CSV (coma delimitada)**” son dos tipos de archivos que se usan en ocasiones para compartir datos de Excel con otros programas.

Después de guardar el trabajo, considere guardar el archivo una segunda vez, con un nombre diferente, para crear una copia de respaldo. Los archivos de discos en los que no se escribe, como los del CD-ROM que acompaña este texto, no se pueden guardar en sus carpetas originales.

### Impresión de las hojas de trabajo

Para imprimir de forma rápida las hojas de trabajo, seleccione **Archivo** → **Imprimir** y después dé clic en el botón **Aceptar** de la ventana de diálogo de imprimir. Sin embargo, salvo por las hojas de cálculo más simples, probablemente deseará tener una

vista preliminar de lo que imprimirá y hacer los ajustes que considere necesarios antes de imprimir. Para hacerlo, seleccione la hoja de trabajo que va a imprimir y después dé clic en **Archivo** → **Vista previa de la impresión**. Si la vista preliminar contiene errores de formato, dé clic en el botón **Cerrar**, haga los cambios necesarios y vuelva a seleccionar **Archivo** → **Vista previa de la impresión**. Cuando esté satisfecho de los resultados, dé clic en el botón **Imprimir** en la ventana de vista preliminar de la impresión y después en el botón **Aceptar** en la ventana de diálogo de Impresión.

La ventana de diálogo de Impresión (vea la figura A1.2 en la página 14) contiene la configuración para seleccionar la impresora que usará, las partes de la hoja de trabajo que imprimirá (la hoja de trabajo activa es por default) y el número de copias que hará (por default es 1). Si requiere cambiar esta configuración, hágalo antes de dar clic en el botón de **Aceptar**.

Después de la impresión, tal vez desee verificar los contenidos. La mayoría de las fallas de impresión muestran un mensaje de error que al mismo tiempo le ayudará a corregirlas. Se personalizan las impresiones en el **Archivo** → **Colocación de hoja** (o dando clic en el botón de Setup en la ventana Vista preliminar de la Impresión) y realizando las entradas apropiados en la ventana de diálogo de Colocación de hoja (no mostrada), antes de imprimir las hojas de trabajo.

## A1.3 INTRODUCCIÓN A MINITAB

### Vista general de Minitab

Minitab es un programa que resultó de los esfuerzos de la Pennsylvania State University para mejorar la enseñanza de la estadística. Actualmente, mientras aún lo utilizan en muchas escuelas, Minitab se ha convertido en un producto comercial que también emplean grandes corporaciones alrededor del mundo, como Ford Motor Company, 3M y GE.

Con Minitab se crean y se abren **proyectos** para guardar los datos y resultados. Los componentes del proyecto son una **sesión** o bitácora de actividades, un **Administrador de proyecto** que resume el contenido del proyecto y cualquier hoja de trabajo o gráficos. Los componentes del trabajo se muestran en ventanas separadas *dentro* de la aplicación de la ventana de Minitab. (Es posible traer cualquier ventana al frente seleccionándola en el menú de Minitab.) Es posible abrir o cerrar un proyecto completo o, como se hace en este texto, abrir y guardar hojas de trabajo. La precisión de Minitab, su disponibilidad para diferentes tipos de sistemas computacionales y la aceptación comercial hacen de este programa una gran herramienta para el aprendizaje de la estadística.

## Uso de las hojas de cálculo de Minitab

Usted ingresa los datos a la hoja de trabajo de Minitab de modo que cada variable es asignada a una columna. Las hojas de trabajo de Minitab están organizadas con base en filas y columnas numeradas en la forma  $C_n$ , donde C1 es la primera columna. Se ingresan los rótulos de las variables en una fila especial sin número que antecede a la fila 1. A diferencia de las hojas de trabajo de otros programas como Excel, las hojas de trabajo de Minitab no aceptan fórmulas y no se recalculan de forma automática cuando se modifican los valores de los datos.

Por default, Minitab abre las hojas de cálculo con nombres seriados en la forma de Hoja de trabajo1, Hoja de trabajo2, etcétera. El mejor nombre será el que refleje el contenido de la hoja de trabajo, como Fondos para una hoja de trabajo que contenga datos sobre los fondos mutuos. Para dar a la hoja de trabajo un nombre descriptivo, abra la ventana de Administrador del proyecto, dé un clic con el botón derecho del ratón en el ícono de la hoja de trabajo y seleccione **Renombrar** en el menú pequeño, donde podrá escribir el nuevo nombre.

## Abrir y guardar las hojas de trabajo y otros componentes

Las hojas de trabajo se abren para usar datos que usted u otros usuarios han creado anteriormente. Para abrir una hoja de trabajo de Minitab, se selecciona primero **Archivo** → **Abrir Hoja de trabajo**. En la ventana de diálogo de Abrir Hoja de trabajo que aparece (vea la figura A1.6), seleccione el archivo que debe abrir y dé clic en el botón **OK**. Si no encuentra su archivo, deberá dar uno o más de los siguientes pasos:

- Use las barras de desplazamiento, si aparecen, para mostrar la lista completa de archivos.
- Seleccione el folder correcto de la lista hacia abajo **Buscar en** que aparece en la parte superior de la ventana de diálogo.
- Cambie el valor de **Tipo de archivos** en la lista hacia abajo en el botón de la ventana de diálogo. Seleccione **Archivos de texto** de la lista para ver cualquier archivo de texto; para listar cada archivo del folder, seleccione **Todos los archivos**.

Para abrir el Proyecto Minitab que pueda incluir la sesión, hoja de trabajo y gráficos, seleccione **Archivo** → **Abrir Proyecto**.

Para guardar una hoja de trabajo, seleccione **Archivo** → **Guardar Hoja de trabajo actual como** para mostrar la ventana de diálogo de Guardar hoja de trabajo como, que es similar a la ventana de diálogo de Abrir hoja de trabajo como. Ingrese (o edite) el nombre del archivo en el cuadro **Nombre del archivo** y dé clic en el botón **OK**. Si es aplicable, también puede hacer lo siguiente:

- Cambie a otro folder seleccionándolo de la lista hacia abajo **Guardar en**.
- Cambie el valor de **Guardar como escrito** a algo más que la opción por default de Minitab, “Minitab Portable” o una versión anterior de Minitab, como “Minitab 13”, que son opciones que se utilizan comúnmente.



**FIGURA A1.6** Abrir la ventana de diálogo de la hoja de trabajo.

Después de guardar el trabajo, considere guardar el archivo una segunda vez, con un nombre diferente, para crear un respaldo de su trabajo. Los archivos que se abren de discos en los que no se puede escribir, como el CD-ROM que acompaña este texto, no se pueden guardar en sus carpetas originales.

Para guardar un Proyecto de Minitab, seleccione también **Archivo** → **Guardar Proyecto como**. La ventana de diálogo de Guardar el Proyecto como incluye el botón **Opciones** que muestra una ventana de diálogo en la que se pueden seleccionar partes del proyecto además de las hojas de cálculo que guardará.

Los gráficos individuales y la sesión también se guardan de forma separada seleccionando primero sus ventanas y después nuevamente **Archivo** → **Guardar Gráfico como** o **Archivo** → **Guardar sesión como**, según sea apropiado. Los gráficos de Minitab pueden guardarse en el formato de gráficos de Minitab o en cualquiera de los formatos de gráficos comunes, y los archivos de Sesiones pueden guardarse como archivos de texto simples o formateados.

## Impresión de hojas de trabajo, gráficos y sesiones

Para imprimir una hoja de trabajo, gráfico o sesión específicos, primero seleccione la ventana de la hoja de trabajo, gráfico o sesión que va a imprimir. Después seleccione **Archivo** → **Imprimir objeto**, donde **objeto** es una **Ventana de hoja de trabajo**, **Gráfico** o **Sesión**, dependiendo de la ventana seleccionada.

Si imprime una gráfica o una ventana de sesión, podrá ver la ventana de diálogo de Imprimir. Si imprime una hoja de trabajo, primero verá un cuadro de diálogo en la Ventana de Datos



**FIGURA A1.7** Cuadro de diálogo de la ventana de datos opciones de impresión.

Opciones de Impresión (figura A1.7), que permite seleccionar formalmente las opciones para la impresión (las que aparecen por default funcionan bien para la mayoría de las impresiones). Dé clic en el botón **OK** para continuar con la ventana de diálogo de Imprimir.

La ventana de diálogo de Imprimir permite seleccionar la impresora que se va a utilizar, qué páginas se van a imprimir y el número de copias que se desea (1 es por default). Si usted desea modificar estas opciones, hágalo antes de dar clic en el botón **OK** para realizar la impresión.

Después de la impresión, deberá verificar su contenido. La mayoría de los errores de impresión generan en la pantalla información que permite conocer la razón de la falla. Es posible cambiar el tamaño o la orientación del papel en la impresión seleccionando **Archivo → Colocación del papel** y hacer las selección apropiada dando clic en el botón **OK**.



# CAPÍTULO 2

## Presentación de datos en tablas y gráficas

---

**USO DE LA ESTADÍSTICA:** Comparación de los rendimientos de los fondos de inversión

### 2.1 TABLAS Y GRÁFICAS PARA DATOS CATEGÓRICOS

Tabla de resumen  
Gráfica de barras  
Gráfica de pastel  
Diagrama de Pareto

### 2.2 ORGANIZACIÓN DE LOS DATOS NUMÉRICOS

Arreglo ordenado  
Diagrama de tallo y hojas

### 2.3 TABLAS Y GRÁFICAS PARA DATOS NUMÉRICOS

Distribución de frecuencias  
Distribución de frecuencias relativas  
y distribución de porcentajes  
Distribución acumulativa  
Histograma

Polígono

Polígono de porcentaje acumulado (ojiva)

### 2.4 TABLAS Y GRÁFICAS DE DATOS BIVARIADOS

Tabla de contingencia  
Gráfica de barras agrupadas

### 2.5 DIAGRAMAS DE DISPERSIÓN Y SERIES DE TIEMPO

Diagrama de dispersión  
Series de tiempo

### 2.6 USO INADECUADO DE GRÁFICAS Y CONSIDERACIONES ÉTICAS

### A.2 USO DEL SOFTWARE PARA TABLAS Y GRÁFICAS

*A2.1 Excel*  
*A2.2 Minitab*  
*A2.3 (Tema CD-ROM) SPSS*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- A desarrollar tablas y gráficas para datos categóricos
- A desarrollar tablas y gráficas para datos numéricos
- Los principios para presentar gráficas de forma adecuada

# USO DE LA ESTADÍSTICA



## Comparación de los rendimientos de los fondos de inversión

Entre las muchas opciones de inversión disponibles en la actualidad, una elección común para quienes piensan en su retiro son los fondos de inversión. Si usted decide invertir en fondos de inversión su cuenta de retiro, ¿qué haría para hacer una elección razonable de entre todos los fondos disponibles hoy?

Primero debería conocer las diferentes categorías de los fondos de inversión. Debería conocer las estrategias de los profesionales que administran los fondos. ¿Invierten en valores de alto riesgo o hacen elecciones más conservadoras? ¿El fondo se especializa en un determinado tamaño de compañía, uno cuya reserva principal totaliza un gran capital o uno de capital reducido? ¿Cobra el fondo comisiones por administración que reducen el porcentaje de utilidad del inversionista? Y, por supuesto, debería conocer qué tan bien el fondo ha manejado las inversiones en el pasado.

Todos éstos son datos que debe revisar cuando considere varias posibilidades de invertir en fondos de inversión. ¿Cómo “poner manos a la obra” con estos datos y explorarlos de manera exhaustiva?

Una de las formas en las que podría responder a las preguntas de “Uso de la estadística” es estudiando los datos de los fondos de inversión. El CD-ROM incluye una muestra de 121 fondos de inversión en el archivo **MUTUALFUNDS2004**. Como inversionista, le gustaría examinar tanto las variables categóricas como las numéricas. ¿Tienen los fondos de inversión como un objetivo de crecimiento rendimientos menores que los fondos de inversión con un objetivo de valor? ¿Tienen los fondos de crecimiento a ser inversiones más riesgosas que los fondos de valor? Este capítulo le ayudará a seleccionar y desarrollar tablas y gráficas apropiadas para encontrar la respuesta a éstas y otras preguntas.

## 2.1 TABLAS Y GRÁFICAS PARA DATOS CATEGÓRICOS

Cuando dispone de datos categóricos usted registra las respuestas en categorías y luego presenta la frecuencia o porcentaje de cada categoría en tablas y gráficas.

### Tabla de resumen

La **tabla de resumen** indica la frecuencia, cantidad o porcentaje de objetos en un conjunto de categorías para observar las diferencias que hay entre ellas. Una tabla de resumen lista las categorías en una columna y la frecuencia, cantidad o porcentaje en una columna o columnas independientes. La tabla 2.1 ilustra la tabla de resumen basada en una reciente encuesta en la que se preguntó por qué la gente compra los regalos en línea durante las temporadas festivas (*USA Today Snapshots, “Convenience, Shipping Make Online Appealing”, USA Today*, 24 de diciembre, 2003, A1). En la tabla 2.1 se observa que las razones más comunes para comprar en línea son los gastos de envío gratis y la conveniencia, seguidos por la comparación de compra. Muy pocos compran en línea por su variedad en la selección o velocidad.

**TABLA 2.1**

Razones por las que se compran regalos en línea durante la temporada festiva.

Razón	Porcentaje
Comparación de compra	23
Conveniencia	33
Gastos de envío gratis	34
Variedad en la selección	6
Velocidad	4
Total	100

**EJEMPLO 2.1****TABLA DE RESUMEN DE LOS NIVELES DE RIESGO DE LOS FONDOS DE INVERSIÓN**

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 22) están clasificados según su nivel de riesgo: bajo, promedio y alto. Construya una tabla de resumen de los fondos de inversión categorizados por el nivel de riesgo.

**SOLUCIÓN**

La mayoría de los fondos de inversión son de riesgo bajo o riesgo promedio (104 o aproximadamente el 86%). Muy pocos de los fondos de inversión son de alto riesgo (14%).

**TABLA 2.2**

Tabla de resumen de frecuencia y porcentaje perteneciente al nivel de riesgo para 121 fondos de inversión.

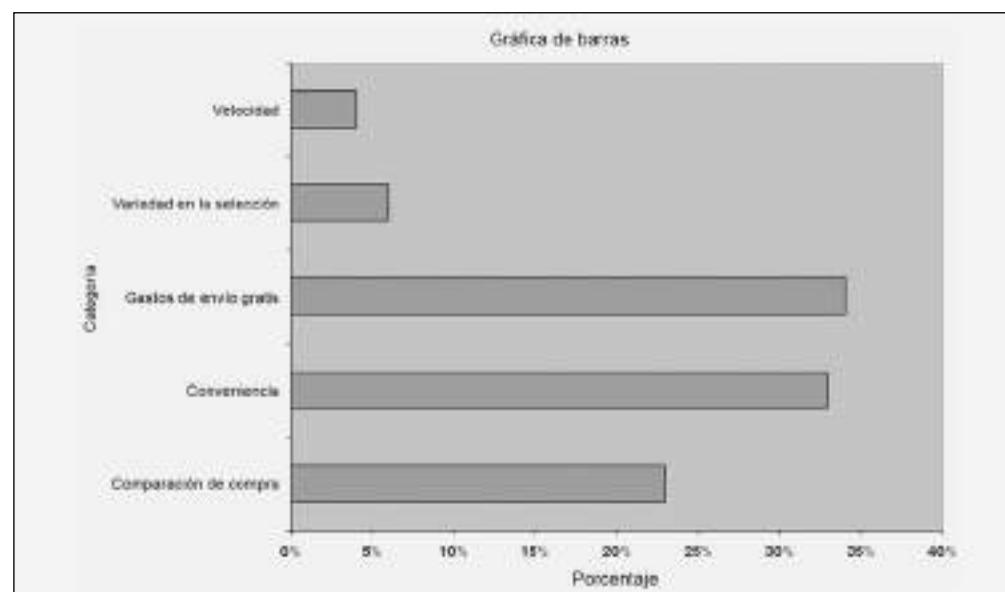
Nivel de riesgo del fondo	Número de fondos	Porcentaje de fondos
Bajo	58	47.93
Promedio	46	38.02
Alto	17	14.05
Total	121	100.00

**Gráfica de barras**

En una **gráfica de barras**, cada barra muestra una categoría, su longitud representa la cantidad, frecuencia o porcentaje de los valores que caen en cada categoría. La figura 2.1 muestra una gráfica de barras para las razones de comprar regalos en línea durante la temporada festiva, de acuerdo con la tabla 2.1.

**FIGURA 2.1**

Gráfica de barras de Excel con las razones por las cuales se compran regalos en línea durante la temporada festiva.



Las gráficas de barras permiten comparar los porcentajes de diferentes categorías. En la figura 2.1 las razones más comunes para comprar en línea son los gastos de envío gratis y la conveniencia, seguidos por la comparación de compra. Muy pocos respondieron que compran en línea por su variedad en la selección o velocidad.

**EJEMPLO 2.2****GRÁFICA DE BARRAS PARA LOS NIVELES DE RIESGO DE LOS FONDOS DE INVERSIÓN**

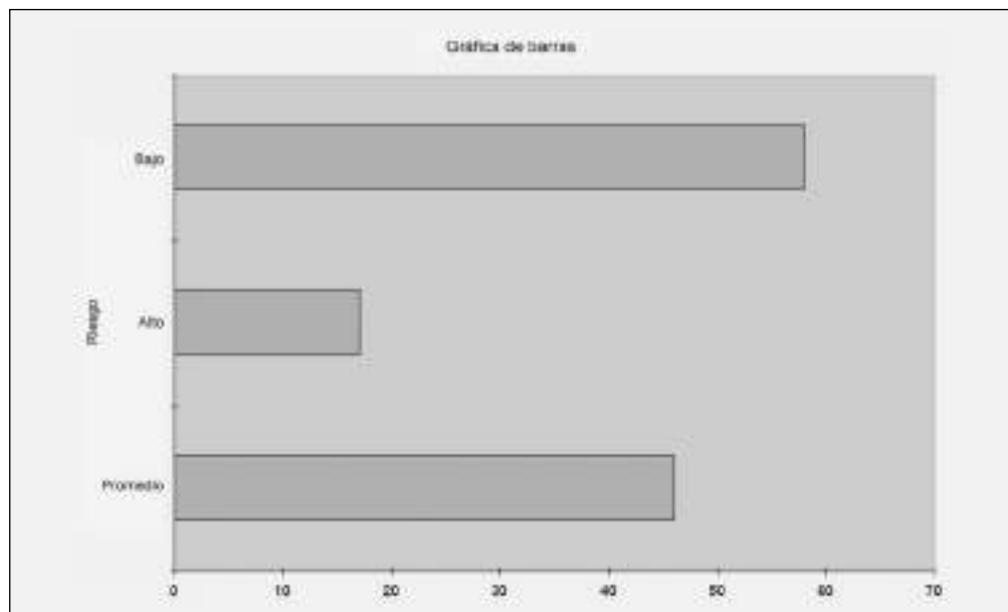
Construya una gráfica de barras para los niveles de riesgo de los fondos de inversión (con base en la información de la tabla 2.2) e interprete los resultados.

**SOLUCIÓN**

La mayoría de los fondos de inversión son de bajo riesgo o de riesgo promedio (104 o el 86%). Muy pocos de los fondos de inversión son de alto nivel de riesgo (17 o el 14%).

**FIGURA 2.2**

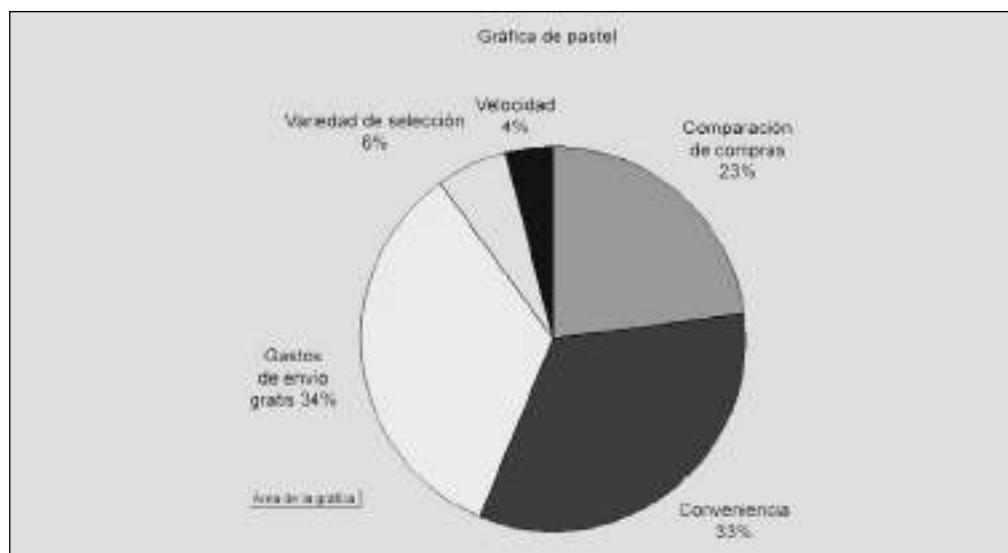
Gráfica de barras de Excel con los niveles de riesgo de los fondos de inversión.

**Gráfica de pastel**

La **gráfica de pastel** es un círculo que se divide en partes para representar las categorías. El tamaño de cada rebanada varía de acuerdo con el porcentaje de cada categoría. En la tabla 2.1, por ejemplo, el 33% de los encuestados afirmaron que la conveniencia era la principal razón para comprar en línea. Así, al construir la gráfica de pastel, los  $360^\circ$  que conforman el círculo se multiplican por 0.33, de lo que resulta una rebanada del pastel que abarca  $118.8^\circ$  de los  $360^\circ$  del círculo. En la figura 2.3 se observa que la gráfica permite visualizar la porción del pastel entero que está en cada categoría. En esta figura, la razón que se refiere a la conveniencia abarca el 33% del pastel y la de la velocidad sólo el 4%.

**FIGURA 2.3**

Gráfica de pastel de Excel con las razones por las que se compran regalos en línea durante la temporada festiva.



¿Qué gráfica se debe utilizar? La selección de una gráfica en particular depende de la intención de quien la construye. Si la comparación de categorías es lo más importante, podría utilizar una gráfica de barras. Si lo importante es observar la parte del total que está en una categoría en particular, debería utilizar la gráfica de pastel.

### EJEMPLO 2.3

#### GRÁFICA DE PASTEL DE LOS NIVELES DE RIESGO DE LOS FONDOS DE INVERSIÓN

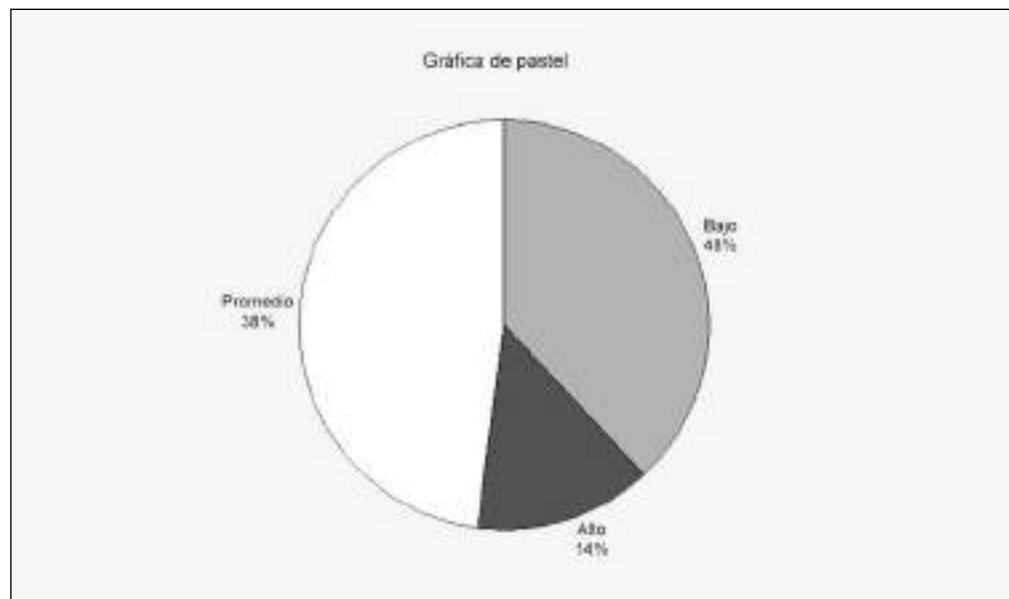
Construya una gráfica de pastel para los niveles de riesgo de los fondos de inversión (vea la tabla 2.2 en la página 23) e interprete los resultados.

#### SOLUCIÓN

(Vea la figura 2.4.) La mayoría de los fondos de inversión son de riesgo bajo o de riesgo promedio (aproximadamente el 86%). Muy pocos son fondos de alto riesgo (aproximadamente el 14%).

### FIGURA 2.4

Gráfica de pastel de Excel con los niveles de riesgo de los fondos de inversión.



### Diagrama de Pareto

En un **diagrama de Pareto** las respuestas categorizadas se trazan en orden descendente de acuerdo con sus frecuencias y se combinan con la línea de porcentaje acumulado en la misma gráfica. Este diagrama permite identificar situaciones en las que se da el principio de Pareto.

#### PRINCIPIO DE PARETO

El **principio de Pareto** existe cuando la mayoría de los elementos de un conjunto de datos caen en un pequeño número de categorías, y las pocas observaciones restantes se dispersan en un gran número de categorías. A menudo nos referimos a estos dos grupos como lo “poco vital” y lo “mucho trivial”.

El diagrama de Pareto permite separar a lo “poco vital” de lo “mucho trivial”, lo que nos permite enfocarnos en las categorías importantes. En las situaciones en las que los datos en estudio consisten en información defectuosa o incompleta, el diagrama de Pareto se convierte en una herramienta valiosa para dar prioridad a los esfuerzos de mejoramiento.

La tabla 2.3 presenta datos de una gran compañía de moldeado de inyección que produce componentes moldeados de plástico para teclados de computadora, lavadoras, automóviles y televisores. Los datos presentados en la tabla 2.3 consisten en todos los teclados de computadora defectuosos producidos durante un periodo de tres meses. **TECLADO**

**TABLA 2.3**

Tabla de resumen de las causas de los defectos en los teclados de computadora en un periodo de tres meses.

Causa	Frecuencia	Porcentaje
Mancha negra	413	6.53
Daño	1,039	16.43
Embarque	258	4.08
Marcas de clavijas	834	13.19
Rasguños	442	6.99
Impacto en el molde	275	4.35
Raya plateada	413	6.53
Marca de hundimiento	371	5.87
Marca de spray	292	4.62
Deformación	1,987	31.42
Total	6,324	100.01*

\*Los resultados difieren ligeramente de los 100.00 por el redondeo.

Fuente: U.H. Acharya y C. Mahesh, "Winning Back the Customer's Confidence: A Case Study on the Application of Design of Experiments to an Injection-Molding Process", Quality Engineering, 11, 1999, 357-363.

La tabla 2.4 presenta un resumen para los defectos de los teclados de computadora, en la que las categorías están ordenadas de acuerdo con el porcentaje (y no alfabéticamente). Los porcentajes acumulados para las categorías ordenadas también forman parte de la tabla.

**TABLA 2.4**

Tabla ordenada de resumen de las causas de los defectos en los teclados de computadora en un periodo de tres meses.

Causa	Frecuencia	Porcentaje	Porcentaje acumulativo
Deformación	1,987	31.42	31.42
Daño	1,039	16.43	47.85
Marca de clavijas	834	13.19	61.04
Rasguños	442	6.99	68.03
Mancha negra	413	6.53	74.56
Raya plateada	413	6.53	81.09
Marca de hundimiento	371	5.87	86.96
Marca de spray	292	4.62	91.58
Impacto en el molde	275	4.35	95.93
Embarque	258	4.08	100.00
Total	6,324	100.01*	

\*Los resultados difieren ligeramente de 100.00 por el redondeo.

En la tabla 2.4 la primera categoría en la lista es la deformación (con el 31.42% de los defectos), seguida por daño (16.43%) y marca de clavijas (13.19%). Las dos categorías más frecuentes —deformación y daño— abarcan el 47.85% de los defectos; las tres categorías más frecuentes —deformación, daño y marcas de clavijas— abarcan el 61.04% de los defectos, y así sucesivamente. La figura 2.5 es un diagrama de Pareto basado en los resultados mostrados en forma tabular en la tabla 2.4.

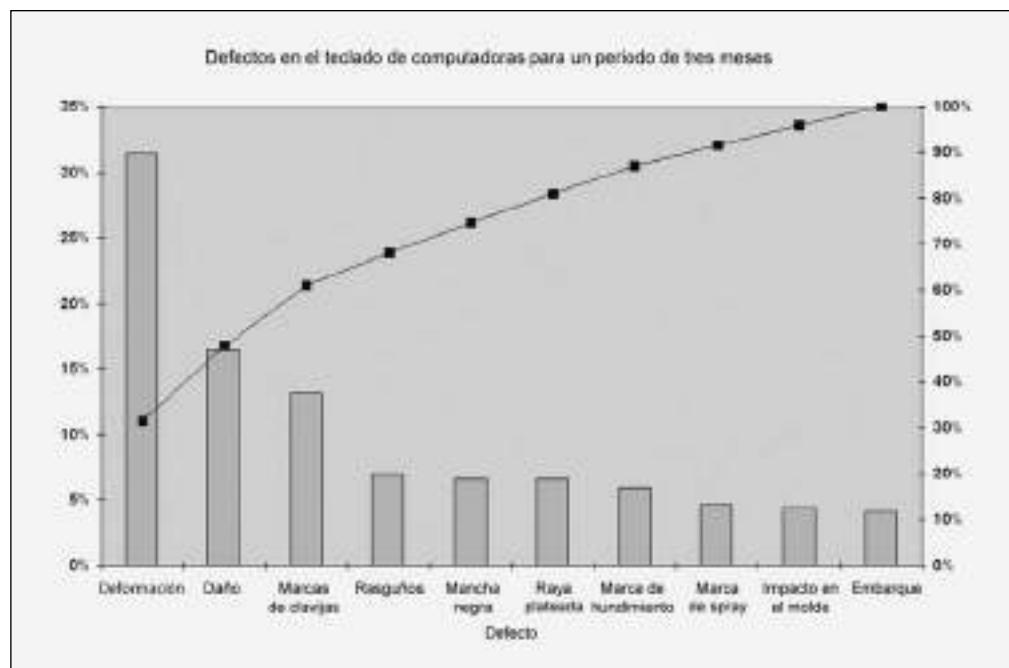
La figura 2.5 presenta las barras de forma vertical a lo largo de la línea de porcentaje acumulado.<sup>1</sup> La línea acumulativa está trazada en el punto medio de cada barra a una altura semejante al porcentaje acumulado. Si sigue la línea, verá que estas tres primeras categorías abarcan más del 60% de las correcciones. Como las categorías del diagrama de Pareto están ordenadas por la frecuencia de ocurrencia, quienes toman decisiones podrán ver dónde concentrar sus esfuerzos para mejorar el proceso. Los intentos de reducir los defectos por deformación, daño y marcas de clavijas deberán generar el mayor gasto. Despues podrán hacerse esfuerzos para reducir los rasguños y las manchas negras.

Para que una tabla de resumen incluya todas las categorías, aun aquellas con menos defectos, en algunos casos se deberá agregar la categoría de *Otros o Miscelánea*. Para estos casos, la barra se coloca a la derecha de las demás.

<sup>1</sup> Lea los porcentajes de la barra en la escala vertical izquierda. Lea los porcentajes acumulados en la escala vertical derecha.

**FIGURA 2.5**

Diagrama de Pareto en Excel para los datos de los defectos de teclado.

**EJEMPLO 2.4****DIAGRAMA DE PARETO DE LAS RAZONES PARA COMPRAR REGALOS EN LÍNEA DURANTE LA TEMPORADA FESTIVA**

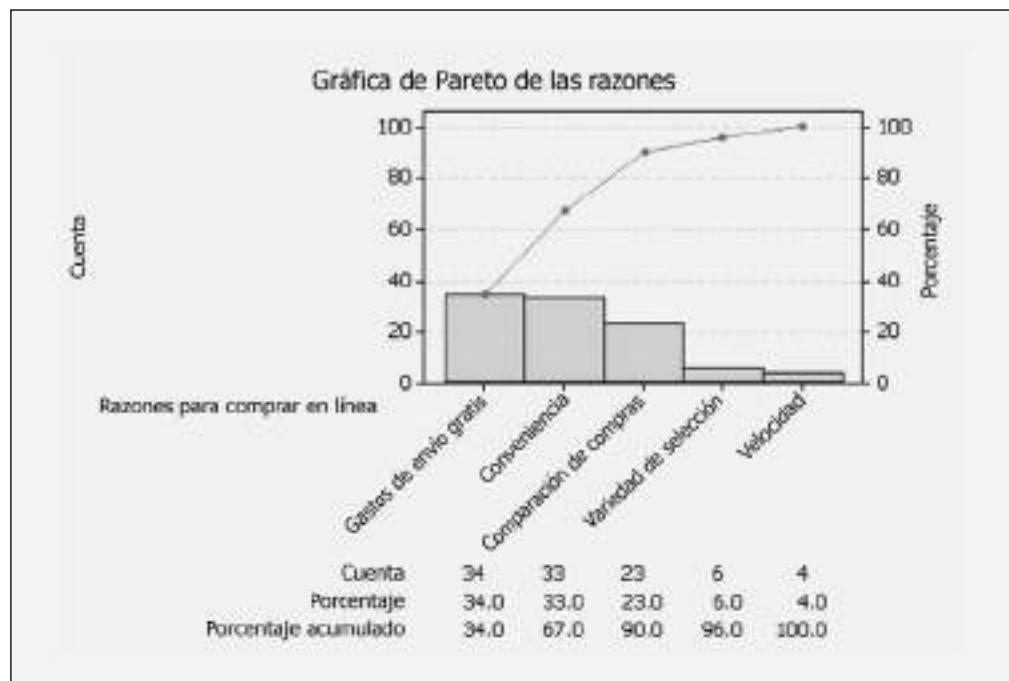
Construya un diagrama de Pareto para las razones de comprar regalos en línea (vea la tabla 2.1 en la página 22).

**SOLUCIÓN**

En la figura 2.6, los gastos de envío gratis y la conveniencia abarcan el 67% de las razones para comprar en línea, mientras que los gastos de envío gratis, conveniencia y comparación de compras abarcan el 90%.

**FIGURA 2.6**

Diagrama de Pareto Minitab, de las razones para comprar regalos en línea durante la temporada festiva.



## PROBLEMAS PARA LA SECCIÓN 2.1

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 2.1** Una variable categórica tiene tres categorías con las siguientes frecuencias:

Categoría	Frecuencia
A	13
B	28
C	9

- a. Calcule el porcentaje de valores en cada categoría.  
 b. Construya una gráfica de barras.  
 c. Construya una gráfica de pastel.  
 d. Construya un diagrama de Pareto.

ASISTENCIA  
de PH Grade

- 2.2** Una variable categórica tiene cuatro categorías con los siguientes porcentajes:

Categoría	Porcentaje	Categoría	Porcentaje
A	12	C	35
B	29	D	24

- a. Construya una gráfica de barras.  
 b. Construya una gráfica de pastel.  
 c. Construya un diagrama de Pareto.

### Aplicación de conceptos

Puede resolver los problemas 2.3 a 2.10 manualmente o usando Excel, Minitab o SPSS.

 **AUTO**  
Examen

- 2.3** En una encuesta se preguntó a 150 ejecutivos cuál creían que era el error más común de los candidatos durante las entrevistas de trabajo. Los resultados (*USA Today Snapshots*, 19 de noviembre, 2001) fueron los siguientes:

Razón	Porcentaje
Poco o nulo conocimiento de la compañía	44
Sin preparación para discutir sus planes profesionales	23
Escaso entusiasmo	16
Falta de contacto visual	5
Sin preparación para discutir sus habilidades/ experiencias	3
Otras razones	9

- a. Construya una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.  
 b. ¿Cuál es el método gráfico que mejor refleja los datos?  
 c. Si fuera un candidato en una entrevista de trabajo, ¿qué errores trataría de evitar especialmente?

- 2.4** Un artículo (M. Mangalindan, N. Wingfield y R. Guth, “Rising Clout of Google Prompts Rush by Internet Rivals to Adapt”, *The Wall Street Journal*, 16 de julio, 2003, A1, A6) analizó la amplia influencia que Google tuvo en Internet a nivel mundial. La siguiente tabla indica la participación de mercado

de los buscadores de la Web, entre los usuarios de Internet de Estados Unidos; el estudio se realizó en mayo de 2003.

Fuente	Porcentaje
Ask Jeeves	3
AOL Time Warner	19
Google	32
MSN-Microsoft	15
Yahoo	25
Otros	6

- a. Elabore una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.  
 b. ¿Cuál método gráfico refleja mejor los datos?  
 c. ¿Qué conclusiones se obtienen respecto a la participación de mercado de los buscadores Web en mayo de 2003?

- 2.5** Los estadounidenses pagaron más de 50 mil millones de dólares en transacciones en línea con tarjetas de crédito en el año 2000 (Byron Acohido, “Microsoft, Banks Battle to Control Your e-info”, *USA Today*, 13 de agosto, 2001, 1B-2B). Estas transacciones se distribuyeron de la siguiente manera:

Tarjeta de crédito	Cantidad (miles de millones de \$)	Porcentaje
American Express	8.04	15.6
Discover	1.97	3.8
MasterCard	15.57	30.2
Visa	25.96	50.4

- a. Construya una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.  
 b. ¿Cuál de los métodos gráficos refleja mejor los datos?

- 2.6** La siguiente tabla representa las fuentes de energía eléctrica utilizadas en Estados Unidos en un año reciente:

Fuente	Porcentaje
Carbón	51
Plantas hidroeléctricas	6
Gas natural	16
Nuclear	21
Petróleo	3
Otras	3

Fuente: Departamento de Energía de Estados Unidos.

- a. Elabore un diagrama de Pareto.  
 b. ¿Qué porcentaje de electricidad se deriva de cualquiera de las siguientes fuentes: carbón, energía nuclear o gas natural?  
 c. Construya una gráfica de pastel.  
 d. ¿Qué gráfica prefiere usar: el diagrama de Pareto o la gráfica de pastel? ¿Por qué?

- 2.7** Un artículo (P. Kitchen, “Retirement Plan: To Keep Working”, *Newsday*, 24 de septiembre, 2003) expuso los resultados de una muestra de 2,001 estadounidenses de entre 50 y 70 años

de edad que tenían empleos de tiempo completo o de medio tiempo. La siguiente tabla representa sus planes de retiro.

Planes	Porcentaje
No tener un trabajo asalariado	29
Iniciar un negocio propio	10
Trabajar tiempo completo	7
Trabajar medio tiempo	46
No sabe	3
Otros	5

- a. Elabore una gráfica de barras y una gráfica de pastel.  
 b. ¿Cuál método gráfico cree que describe mejor los datos?

**2.8** El correo electrónico basura (spam) se ha convertido en un problema muy grave para la productividad (J. Hopkins, “Spam Blaster Does Job for Merril”, *USA Today*, 7 de enero, 2004). La siguiente tabla muestra el uso que una compañía da al software antispam con base en una encuesta realizada a ejecutivos de tecnología.

Uso de software antispam por la compañía	Porcentaje
Tiene software para algunos usuarios	12
Tiene software para todos los usuarios	59
Planea tener software en los próximos 12 meses	20
No planea tener software	9

- a. Construya una gráfica de barras y una gráfica de pastel.  
 b. ¿Cuál método gráfico representa mejor estos datos?

**2.9** Un analista registró las causas de las caídas de una red durante los pasados seis meses.

Razón de la falla	Frecuencia
Conexión física	1
Falla de energía	3
Software del servidor	29
Hardware del servidor	2
Servidor falto de memoria	32
Banda ancha inadecuada	1

- a. Realice un diagrama de Pareto.  
 b. Analice las razones “poco vital” y “muy trivial” por las que el sistema de red se cayó.

**2.10** Los siguientes datos representan las quejas acerca de las habitaciones de un hotel.

Razón	Número
Habitación sucia	32
Habitación sin equipamiento	17
Habitación que no está lista	12
Habitación demasiado ruidosa	10
Habitación que requiere de mantenimiento	17
Habitación con pocas camas	9
Habitación que no tiene las características prometidas	7
No tiene instalaciones especiales	2

- a. Elabore un diagrama de Pareto.  
 b. ¿En qué se debe enfocar el hotel si desea reducir el número de quejas? Explique su respuesta.

## 2.2 ORGANIZACIÓN DE LOS DATOS NUMÉRICOS

Cuando el número de datos es grande, es conveniente organizar los datos numéricos en un arreglo ordenado o diagrama de tallo y hojas para ayudar a comprender la información. Suponga que decide llevar a cabo un estudio comparativo del costo de una comida en un restaurante de una gran ciudad con el de una comida similar en un restaurante fuera de la ciudad. La tabla 2.5 muestra los datos de 50 restaurantes citadinos y 50 fuera de la ciudad. **RESTRATE** Los datos no están ordenados de menor a mayor. Esta organización hace difícil obtener conclusiones sobre el precio de las comidas en las dos áreas geográficas.

**TABLA 2.5**

Precio por persona en 50 restaurantes citadinos y en 50 restaurantes fuera de la ciudad.

Ciudad									
50	38	43	56	51	36	25	33	41	44
34	39	49	37	40	50	50	35	22	45
44	38	14	44	51	27	44	39	50	35
31	34	48	48	30	42	26	35	32	63
36	38	53	23	39	45	37	31	39	53
Fuera de la ciudad									
37	37	29	38	37	38	39	29	36	38
44	27	24	34	44	23	30	32	25	29
43	31	26	34	23	41	32	30	28	33
26	51	26	48	39	55	24	38	31	30
51	30	27	38	26	28	33	38	32	25

## Arreglo ordenado

Un **arreglo ordenado** es una secuencia de datos ordenados del menor al mayor valor. La tabla 2.6 contiene la formación ordenada para el precio de las comidas en restaurantes citadinos y de los suburbios. **RESTRATE** En la tabla 2.6 se observa que el precio de una comida en los restaurantes citadinos se encuentra entre \$14 y \$63, y que en los restaurantes fuera de la ciudad se encuentra entre \$23 y \$55.

**TABLA 2.6**

Arreglo ordenado del precio por persona en 50 restaurantes de la ciudad y 50 restaurantes fuera de la ciudad.

<b>Ciudad</b>									
14	22	23	25	26	27	30	31	31	32
33	34	34	35	35	35	36	36	37	37
38	38	38	39	39	39	39	40	41	42
43	44	44	44	44	45	45	48	48	49
50	50	50	50	51	51	53	53	56	63
<b>Fuera de la ciudad</b>									
23	23	24	24	25	25	26	26	26	26
27	27	28	28	29	29	29	30	30	30
30	31	31	32	32	32	33	33	34	34
36	37	37	37	38	38	38	38	38	38
39	39	41	43	44	44	48	51	51	55

## Diagrama de tallo y hojas

El **diagrama de tallo y hojas** organiza los datos en grupos (llamados tallos), para que los valores dentro de cada grupo (las hojas) ramifiquen hacia la derecha de cada fila. El diagrama resultante permite ver cómo se distribuyen y dónde están las concentraciones de datos. Para ver cómo se elabora un diagrama de tallo y hojas suponga que 15 alumnos de su clase comen en un restaurante de comida rápida. Los siguientes datos son las cantidades que gastaron.

5.35 4.75 4.30 5.47 4.85 6.62 3.54 4.87 6.26 5.48 7.27 8.45 6.05 4.76 5.91

Para formar un diagrama de tallo y hojas se colocan los primeros valores en orden ascendente. Se utiliza la columna de las unidades como el tallo y se redondean los decimales (las hojas) a un lugar decimal.

3	5
4	83998
5	4559
6	631
7	3
8	5

El primer valor de 5.35 se redondea a 5.4. Su tallo (fila) es 5 y su hoja es 4. El segundo valor de 4.75 se redondea a 4.8. Su tallo (fila) es 4 y su hoja es 8.

### EJEMPLO 2.5

#### DIAGRAMA DE TALLO Y HOJAS DEL RENDIMIENTO DE FONDOS DE INVERSIÓN EN 2003

En el escenario “Uso de la estadística”, se le pide estudiar el rendimiento de los fondos de inversión en 2003 **MUTUALFUNDS2004**. Elabore un diagrama de tallo y hojas.

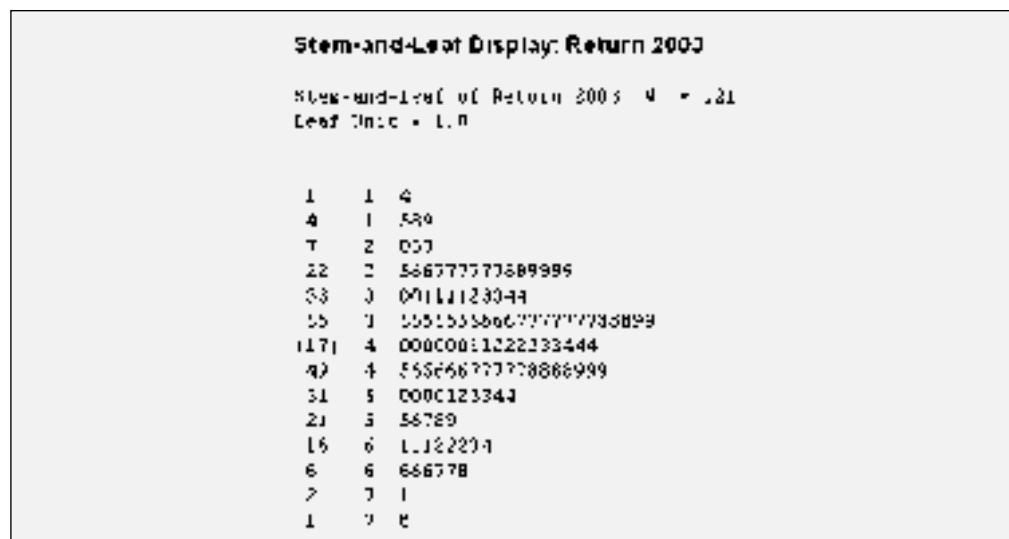
#### SOLUCIÓN

A partir de la figura 2.7 se concluye que:

- El menor rendimiento en 2003 fue del 14%.
- El mayor rendimiento en 2003 fue del 78%.
- Los rendimientos en 2003 se concentraron entre el 25 y 50%.
- Sólo cuatro fondos de inversión dieron rendimientos por debajo del 20% y sólo dos fondos de inversión dieron rendimientos por encima del 70%.

**FIGURA 2.7**

Pantalla de tallo y hojas para los rendimientos en 2003.

**PROBLEMAS PARA LA SECCIÓN 2.2****Aprendizaje básico**

ASISTENCIA  
de PH Grade

- 2.11** Elabore un arreglo ordenado para los siguientes datos obtenidos de una muestra de  $n = 7$  de las puntuaciones parciales de la materia de finanzas:

68 94 63 75 71 88 64

ASISTENCIA  
de PH Grade

- 2.12** Para los siguientes datos obtenidos de una muestra de  $n = 7$  de las puntuaciones parciales de la materia de sistemas de información, realice un diagrama de tallo y hojas:

80 54 69 98 93 53 74

ASISTENCIA  
de PH Grade

- 2.13** Elabore un arreglo ordenado para los siguientes datos obtenidos de una muestra de  $n = 7$  de las puntuaciones parciales de la materia de marketing:

88 78 78 73 91 78 85

ASISTENCIA  
de PH Grade

- 2.14** Elabore un arreglo ordenado a partir del diagrama de tallo y hojas con la siguiente muestra de  $n = 7$  puntuaciones parciales de la materia de sistemas de información:

5	0
6	
7	446
8	19
9	2

**Aplicación de conceptos**

ASISTENCIA  
de PH Grade

- 2.15** El siguiente diagrama de tallo y hojas representa la cantidad de gasolina comprada en galones (con hojas en decenas de galones) para una muestra de 25 autos que utilizan una estación de servicio en la autopista de Nueva Jersey:

9	147
10	02238
11	125566777
12	223489
13	02

- Coloque los datos en un arreglo ordenado.
- ¿Cuál de los dos diagramas aporta más información? Explique su respuesta.
- ¿Cuánta gasolina (en galones) es más probable que se compre?
- ¿Existe una concentración en la compra de cantidades en el centro de distribución?

- ASISTENCIA  
de PH Grade
- 2.16** Los siguientes datos representan las cuotas en dólares de cheques rechazados de una muestra de 23 bancos, firmados por clientes que depositan directamente y que mantienen un saldo promedio de \$100. **BANK-COST1**

26 28 20 20 21 22 25 25 18 25 15 20  
18 20 25 25 22 30 30 30 15 20 29

Fuente: "The New Face of Banking", Copyright © 2000 by Consumers Union of U.S., Inc., Yonkers NY 10703-1057. Adaptado con el permiso de Consumer Reports, junio de 2000.

- Coloque los datos en un arreglo ordenado.
- Elabore un diagrama de tallo y hojas para estos datos.
- ¿Cuál de estos dos diagramas aporta más información? Explique su respuesta.
- ¿Alrededor de qué valor, si lo hay, se encuentran concentradas las cuotas para cheques rechazados? Explique su respuesta.

ASISTENCIA  
de PH Grade

**2.17** Los siguientes datos representan la cuota mensual de servicio que se cobra en dólares a un cliente si su cuenta no alcanza el saldo promedio requerido, en una muestra de clientes de 26 bancos que depositan directamente y que mantienen un saldo promedio de \$1,500. **BANKCOST2**

12	8	5	5	6	6	10	10	9	7	10	7	7
5	0	10	6	9	12	0	5	10	8	5	5	9

Fuente: "The New Face of Banking", Copyright © 2000 by Consumers Union of U.S., Inc., Yonkers NY 10703-1057. Adaptado con el permiso de Consumer Reports, junio de 2000.

- Coloque los datos en un arreglo ordenado.
- Elabore un diagrama de tallo y hojas para estos datos.
- ¿Cuál de estos dos diagramas aporta más información? Explique su respuesta.
- ¿Alrededor de qué valor, si lo hay, se encuentran concentradas las cuotas de servicio? Explique su respuesta.

**AUTO**  
Examen **2.18** Los siguientes datos representan el total de grasa en las hamburguesas y pollo de una muestra de cadenas de comida rápida. **FASTFOOD**

#### HAMBURGUESAS

19	31	34	35	39	39	43
----	----	----	----	----	----	----

#### POLLO

7	9	15	16	16	18	22	25	27	33	39
---	---	----	----	----	----	----	----	----	----	----

Fuente: "Quick Bites", Copyright © 2001 by Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, marzo de 2001.

- Coloque los datos para las hamburguesas y el pollo en dos arreglos ordenados.
- Elabore diagramas de tallo y hojas para las hamburguesas y el pollo.
- ¿Cuál brinda mayor información: el arreglo ordenado o el diagrama de tallo y hojas? Explique.
- Compare las hamburguesas y el pollo en términos de su contenido total de grasa. ¿A qué conclusión llega?

**2.19** Los siguientes datos representan el costo promedio diario de hotel y de la renta de auto para 20 ciudades de Estados Unidos durante una semana en octubre de 2003. **HOTEL-CAR**

Ciudad	Hotel	Autos
San Francisco	205	47
Los Ángeles	179	41
Seattle	185	49
Phoenix	210	38
Denver	128	32
Dallas	145	48
Houston	177	49
Minneapolis	117	41
Chicago	221	56
St. Louis	159	41
Nueva Orleáns	205	50
Detroit	128	32
Cleveland	165	34
Atlanta	180	46
Orlando	198	41
Miami	158	40
Pittsburg	132	39
Boston	283	67
Nueva York	269	69
Washington, D.C.	204	40

Fuente: The Wall Street Journal, 10 de octubre, 2003, W4.

- Coloque los datos para el costo del hotel y el costo de la renta de auto en dos arreglos ordenados.
- Elabore un diagrama de tallo y hojas para el costo del hotel y el costo de la renta de auto.
- ¿Qué aporta mayor información: el arreglo ordenado o el diagrama de tallo y hojas? Explique su respuesta.
- ¿Alrededor de qué valor, si lo hay, se concentran los costos del hotel y de la renta de auto? Explique su respuesta.

## 2.3 TABLAS Y GRÁFICAS PARA DATOS NUMÉRICOS

Cuando tenemos un conjunto de datos muy grande, a menudo se dificulta llegar a conclusiones con base en un arreglo ordenado o en un diagrama de tallo y hojas. En tales circunstancias es necesario usar tablas y gráficas. Existen diferentes tablas y gráficas que permiten presentar visualmente los datos numéricos. Entre ellas se incluyen las distribuciones de frecuencia y de porcentaje, el histograma, el polígono y el polígono de porcentaje acumulado (ojiva).

### Distribución de frecuencias

La distribución de frecuencias nos ayuda a obtener conclusiones de un conjunto de datos grande.

Una **distribución de frecuencias** es una tabla de resumen en la que los datos están organizados en clases o grupos numéricamente ordenados.

Al construir una distribución de frecuencia, se debe prestar atención al seleccionar el *número* apropiado de **agrupaciones o clases** para la tabla, determinando una *amplitud* conveniente de las clases y estableciendo los *límites* de cada una para evitar el traslape.

El número de clases usadas depende del número de valores. Un mayor número de valores permite un mayor número de clases. En general, la distribución de frecuencias debería tener por lo menos cinco clases, pero no más de 15. Tener muy pocas o muchas clases ofrece poca información.

Cuando se elabora una distribución de frecuencias, se define cada clase con base en intervalos de clase de la misma amplitud. Para determinar la **amplitud de un intervalo de clase**, se divide el **rango** (el valor mayor menos el valor menor) de los datos entre el número de los agrupamientos o clases que se desea.

#### DETERMINACIÓN DE LA AMPLITUD DEL INTERVALO DE CLASE

$$\text{Amplitud del intervalo} = \frac{\text{rango}}{\text{número de clases deseado}} \quad (2.1)$$

Los datos de los restaurantes citadinos constituyen una muestra de 50 establecimientos. Para este tamaño de muestra es aceptable tener 10 agrupamientos o clases. En la formación ordenada de la tabla 2.6 en la página 30, el rango de los datos es  $\$63 - \$14 = \$49$ . Mediante la ecuación (2.1), se aproxima la amplitud del intervalo de clase de la siguiente forma:

$$\text{Amplitud del intervalo} = \frac{49}{10} = 4.9$$

Se debe elegir una amplitud de intervalo que simplifique la lectura e interpretación. Por tanto, en lugar de usar una amplitud de intervalo de \$4.90, se debería elegir una amplitud de intervalo de \$5.00.

Para construir una tabla de distribución de frecuencias, se deberán establecer **límites de clase** claramente definidos para cada agrupamiento de clase, para que los valores estén clasificados adecuadamente. Cada valor se coloca en una y sólo una clase. Se debe evitar el traslape de clases.

Puesto que se ha establecido la amplitud de cada intervalo de clase para el costo de la comida en \$5, es necesario establecer los límites para los diferentes agrupamientos o clases, para que así se incluya el rango completo de valores. Siempre que sea posible, uno debería elegir tales límites para simplificar la lectura e interpretación. Así, como los costos varían de \$14 a \$63, para los restaurantes de la ciudad, el primer intervalo de clase va de \$10 a menos de \$15, el segundo va de \$15 a menos de \$20, y así sucesivamente, hasta que se hayan formado 11 clases. Cada clase tiene una amplitud de intervalo de \$5, sin traslaparse. El centro de cada clase, el **punto medio de la clase**, está a la mitad del camino entre el límite inferior y el límite superior de la clase. Por tanto, el punto medio de la clase que va de \$10 a por debajo de \$15 es \$12.5, el punto medio de la clase que va de \$15 a por debajo de \$20 es \$17.5, etcétera. La tabla 2.7 es una distribución de frecuencias para el costo por comida de los 50 restaurantes de la ciudad y para los 50 restaurantes fuera de la ciudad.

**TABLA 2.7**

Distribución de frecuencias del costo por comida para 50 restaurantes de la ciudad y 50 restaurantes de los suburbios.

Costo por comida (\$)	Frecuencia de la ciudad	Frecuencia de los suburbios
10 pero menos de \$15	1	0
15 pero menos de \$20	0	0
20 pero menos de \$25	2	4
25 pero menos de \$30	3	13
30 pero menos de \$35	7	13
35 pero menos de \$40	14	12
40 pero menos de \$45	8	4
45 pero menos de \$50	5	1
50 pero menos de \$55	8	2
55 pero menos de \$60	1	1
60 pero menos de \$65	1	0
Total	50	50

La distribución de frecuencias permite obtener conclusiones acerca de las características principales de los datos. Por ejemplo, la tabla 2.7 muestra que el costo de las comidas en los restaurantes de la ciudad está concentrado entre los \$30 y los \$55, en comparación con las comidas en los restaurantes de los suburbios, los cuales están concentrados entre los \$25 y los \$40.

Si el conjunto de datos no contiene muchos valores, un conjunto de límites de clase refleja una imagen diferente de la que da otro conjunto de límites. Por ejemplo, para los datos del costo del restaurante, usar un intervalo de clase de amplitud 4.0 en lugar de 5.0 (como el que se utilizó en la tabla 2.7), provocaría cambios en la forma en la que los valores se distribuyen entre las clases.

Usted obtendrá cambios en la concentración de los datos al elegir límites de clase inferiores y superiores diferentes. Por fortuna, conforme aumenta el tamaño de la muestra, las alteraciones en la selección de los límites de clase afectan cada vez menos la concentración de los datos.

## EJEMPLO 2.6

### DISTRIBUCIÓN DE FRECUENCIAS DE LOS RENDIMIENTOS EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

En el escenario de “Uso de la estadística” se le pide comparar el rendimiento en 2003 de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Construya una distribución de frecuencias para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

El porcentaje de rendimientos en 2003 de los fondos de crecimiento está concentrado significativamente entre el 30 y el 50, con una ligera concentración entre el 20 y el 30 (vea la tabla 2.8). El porcentaje de rendimientos en 2003 de los fondos de valor está concentrado entre el 30 y el 50, con algunos entre 20 y 30 y entre 50 y 70. No debe comparar directamente las frecuencias de los fondos de crecimiento y los fondos de valor puesto que en la muestra hay 49 fondos de crecimiento y 72 fondos de valor.

**TABLA 2.8**

Distribución de frecuencias del rendimiento en 2003 de los fondos de inversión de crecimiento y de valor.

Porcentaje de rendimiento 2003	Frecuencia de crecimiento	Frecuencia de valor
10 pero menos que 20	2	2
20 pero menos que 30	9	9
30 pero menos que 40	13	20
40 pero menos que 50	15	20
50 pero menos que 60	5	10
60 pero menos que 70	5	9
70 pero menos que 80	0	2
Total	49	72

### Distribución de frecuencias relativas y distribución de porcentajes

Como generalmente deseamos saber la proporción o el porcentaje del total en cada grupo, es preferible usar la distribución de frecuencias relativa o la distribución de porcentajes. Cuando comparamos dos o más grupos que difieren en el tamaño de su muestra, se debe usar una distribución de frecuencias relativa o una distribución de porcentaje.

Se crea una **distribución de frecuencias relativa** al dividir las frecuencias de cada clase de la distribución de frecuencias (vea la tabla 2.7 en la página 33) por el número total de valores. Se crea una **distribución de porcentajes** al multiplicar cada frecuencia relativa por 100%. Así, la frecuencia relativa de las comidas en los restaurantes de la ciudad que cuestan entre \$30 y \$35 es 7 dividido por 50 o 0.14, y el porcentaje es del 14%. La tabla 2.9 presenta la distribución de frecuencias relativa y la distribución de porcentajes del costo de las comidas en restaurantes de la ciudad y de los suburbios.

A partir de la tabla 2.9, se concluye que las comidas cuestan más en los restaurantes de la ciudad que en los de los suburbios: el 16% de las comidas en los restaurantes de la ciudad cuestan entre \$50 y \$55, en comparación con el 4% de los restaurantes de los suburbios; mientras que sólo el 6% de las comidas en los restaurantes de la ciudad cuestan entre \$25 y \$30 en comparación con el 26% de los restaurantes de los suburbios.

**TABLA 2.9**

Distribución de frecuencias relativa y distribución de porcentajes del costo de las comidas en restaurantes de la ciudad y de los suburbios.

Costo por comida (\$)	Ciudad			Suburbios	
	Frecuencia relativa	Porcentaje		Frecuencia relativa	Porcentaje
10 pero menos que \$15	0.02	2.0		0.00	0.0
15 pero menos que \$20	0.00	0.0		0.00	0.0
20 pero menos que \$25	0.04	4.0		0.08	8.0
25 pero menos que \$30	0.06	6.0		0.26	26.0
30 pero menos que \$35	0.14	14.0		0.26	26.0
35 pero menos que \$40	0.28	28.0		0.24	24.0
40 pero menos que \$45	0.16	16.0		0.08	8.0
45 pero menos que \$50	0.10	10.0		0.02	2.0
50 pero menos que \$55	0.16	16.0		0.04	4.0
55 pero menos que \$60	0.02	2.0		0.02	2.0
60 pero menos que \$65	0.02	2.0		0.00	0.0
Total	1.00	100.0		1.00	100.0

**EJEMPLO 2.7**

### DISTRIBUCIÓN DE FRECUENCIAS RELATIVA Y DISTRIBUCIÓN DE PORCENTAJES DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

En el escenario de “Uso de la estadística”, se le pide que compare el rendimiento en 2003 de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Construya una distribución de frecuencias relativa y una distribución de porcentajes para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

Se concluye (vea la tabla 2.10) que el rendimiento en 2003 de los fondos de crecimiento es ligeramente inferior a la de los fondos de valor y que el 18.37% de los fondos de crecimiento tienen rendimientos entre 20 y 30 en comparación con el 12.5% de los fondos de valores. Los fondos de valor tienen rendimientos ligeramente mayores (entre 50 y 60, y entre 60 y 70) que los fondos de crecimiento.

**TABLA 2.10**

Distribución de frecuencias relativa y distribución de porcentajes del rendimiento en 2003 de los fondos de inversión de crecimiento y de valor.

Porcentaje anual de rendimiento en 2003	Crecimiento		Valor	
	Proporción	Porcentaje	Proporción	Porcentaje
10 pero menos que 20	0.0408	4.08	0.0278	2.78
20 pero menos que 30	0.1837	18.37	0.1250	12.50
30 pero menos que 40	0.2653	26.53	0.2778	27.78
40 pero menos que 50	0.3061	30.61	0.2778	27.78
50 pero menos que 60	0.1020	10.20	0.1389	13.89
60 pero menos que 70	0.1020	10.20	0.1250	12.50
70 pero menos que 80	0.0000	0.00	0.0278	2.78
Total	1.0000	100.0	1.0000	100.0

### Distribución acumulativa

La **distribución de porcentaje acumulado** constituye una manera de presentar la información del porcentaje de los valores que están por debajo de cierto valor. Por ejemplo, tal vez se desea conocer qué porcentaje de las comidas de los restaurantes de la ciudad cuestan menos que \$20, menos que \$30, menos que \$50, etcétera. La distribución de porcentaje se usa para formar una distribución de porcentaje acumulado. A partir de la tabla 2.12, se sabe que el 0.00% de las comidas cuestan menos de \$10, el 2% cuesta menos de \$15, el 2% también cuesta menos de \$20 (porque ninguna de las comidas cuestan entre \$15 y \$20), el 6% (2 + 4%) cuesta menos de \$25, y así sucesivamente, hasta que el 100% de comidas cuestan menos de \$65. La tabla 2.11 ilustra cómo desarrollar la distribución de porcentaje acumulado para el costo de las comidas en restaurantes de la ciudad.

**TABLA 2.11**

Desarrollo de la distribución de porcentaje acumulado del costo de las comidas en los restaurantes de la ciudad.

Costo por comida (\$)	Porcentaje	Porcentaje de fondos por debajo del límite inferior del intervalo de clase
10 pero menos de \$15	2	0
15 pero menos de \$20	0	2
20 pero menos de \$25	4	$2 = 2 + 0$
25 pero menos de \$30	6	$6 = 2 + 0 + 4$
30 pero menos de \$35	14	$12 = 2 + 0 + 4 + 6$
35 pero menos de \$40	28	$26 = 2 + 0 + 4 + 6 + 14$
40 pero menos de \$45	16	$54 = 2 + 0 + 4 + 6 + 14 + 28$
45 pero menos de \$50	10	$70 = 2 + 0 + 4 + 6 + 14 + 28 + 16$
50 pero menos de \$55	16	$80 = 2 + 0 + 4 + 6 + 14 + 28 + 16 + 10$
55 pero menos de \$60	2	$96 = 2 + 0 + 4 + 6 + 14 + 28 + 16 + 10 + 16$
60 pero menos de \$65	2	$98 = 2 + 0 + 4 + 6 + 14 + 28 + 16 + 10 + 16 + 2$
\$65 pero menos de \$70	0	$100 = 2 + 0 + 4 + 6 + 14 + 28 + 16 + 10 + 16 + 2 + 2$

La tabla 2.12 resume los porcentajes acumulados del costo de las comidas en restaurantes de la ciudad y de los suburbios. La distribución acumulativa muestra claramente que los costos de la comida son inferiores en los restaurantes de los suburbios que en los de la ciudad: en el 34% de los restaurantes de los suburbios cuesta menos de \$30, en comparación con sólo el 12% de los restaurantes de la ciudad; en el 60% de los restaurantes de los suburbios cuesta menos de \$35 en comparación con sólo el 26% de los restaurantes de la ciudad; en el 84% de los restaurantes de los suburbios cuesta menos de \$40 en comparación con sólo el 54% de los restaurantes de la ciudad.

**TABLA 2.12**

Distribución de porcentajes acumulados del costo de las comidas en restaurantes de la ciudad y de los suburbios.

Costo (\$)	Porcentaje de restaurantes de la ciudad con valor menor al indicado	Porcentaje de restaurantes de los suburbios con valor menor al indicado
10	0	0
15	2	0
20	2	0
25	6	8
30	12	34
35	26	60
40	54	84
45	70	92
50	80	94
55	96	98
60	98	100
65	100	100

### EJEMPLO 2.8

#### DISTRIBUCIÓN DE PORCENTAJE ACUMULADO DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

En el escenario “Uso de la estadística”, se le pide comparar el rendimiento anual en 2003 de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Construya una distribución de porcentaje acumulado para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

La distribución acumulativa de la tabla 2.13 indica que los fondos de crecimiento tienen un rendimiento ligeramente mayor que los fondos de valor: el 22.45% de los fondos de crecimiento tienen rendimientos por debajo de 30 en comparación con el 15.28% de los fondos de valor; el 48.98% de los fondos de crecimiento tienen rendimientos por debajo de 40 en comparación con el 43.06% de los fondos de valor; el 79.59% de los fondos de crecimiento tienen rendimientos por debajo de 50 en comparación con el 70.83% de los fondos de valor.

**TABLA 2.13**

Distribuciones de porcentaje acumulado del rendimiento en 2003 de los fondos de crecimiento y de valor.

Rendimiento anual	Porcentaje menor del valor indicado del fondo de crecimiento	Porcentaje menor del valor indicado del fondo de valor
10	0.00	0.00
20	4.08	2.78
30	22.45	15.28
40	48.98	43.06
50	79.59	70.83
60	89.80	84.72
70	100.00	97.22
80	100.00	100.00

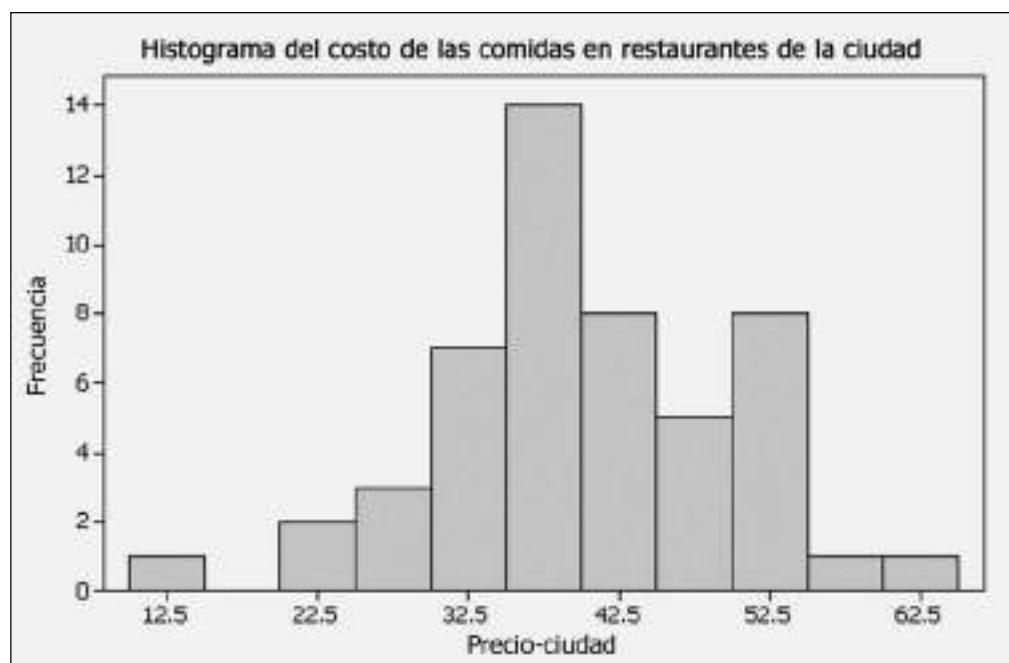
### Histograma

El **histograma** es una gráfica de barras para datos numéricos agrupados en los que las frecuencias o los porcentajes de cada grupo de datos numéricos están representados por barras individuales. En un histograma, no hay brechas entre las barras adyacentes como en la gráfica de barras de los datos categóricos. La variable que nos interesa se coloca a lo largo del eje ( $X$ ) horizontal. El eje ( $Y$ ) vertical representa la frecuencia o el porcentaje de los valores por intervalo de clase.

La figura 2.8 muestra un histograma de frecuencia Minitab para el costo de las comidas en los restaurantes de la ciudad. El histograma indica que el costo de las comidas en los restaurantes de la ciudad se concentra entre aproximadamente \$30 y \$55. Muy pocas comidas cuestan menos de \$20 o más de \$55.

**FIGURA 2.8**

Histograma Minitab para el costo de las comidas en restaurantes de la ciudad.

**EJEMPLO 2.9**

### HISTOGRAMA PARA EL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

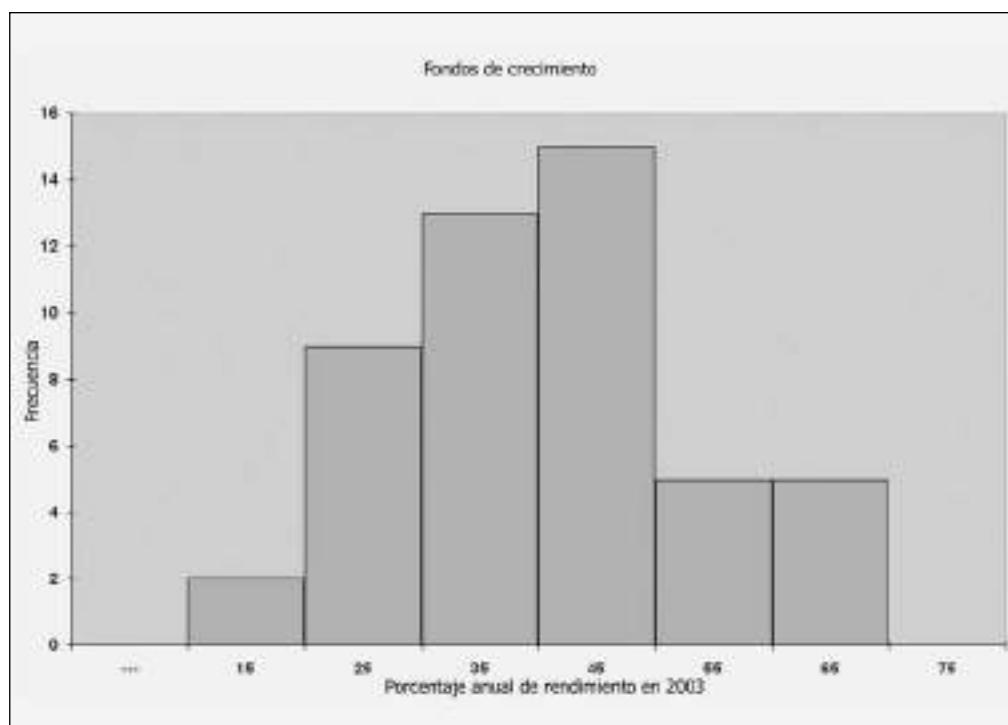
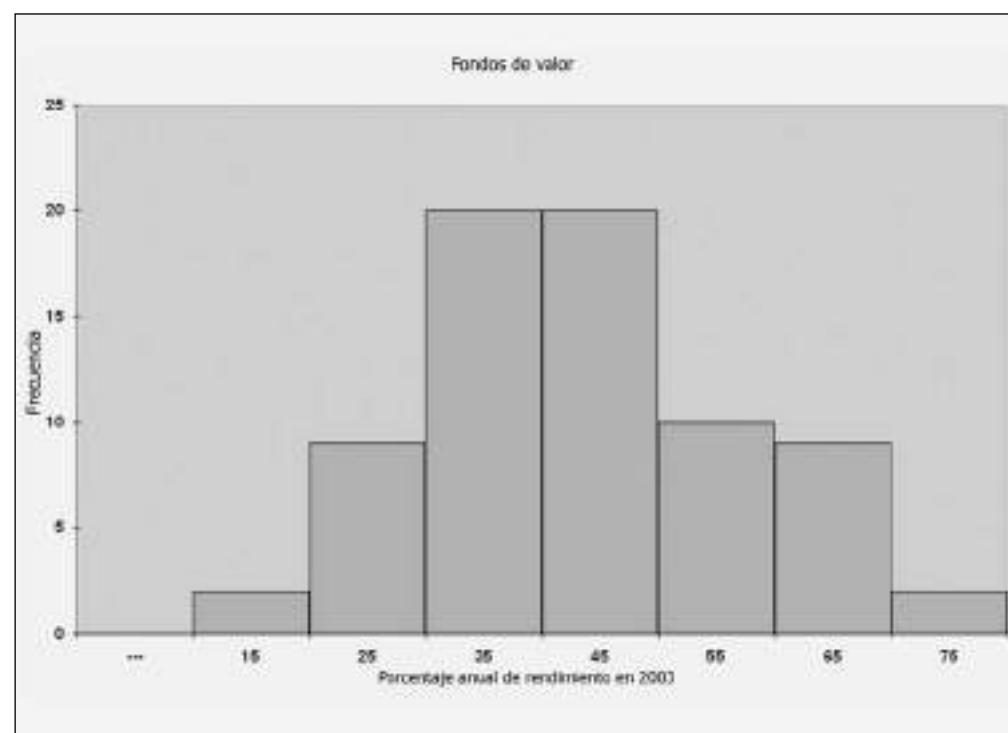
En el escenario “Uso de la estadística” le interesa comparar el rendimiento en 2003 de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Construya histogramas para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

La figura 2.9 muestra que la distribución de los fondos de crecimiento tiene rendimientos menores en comparación con los fondos de valor, los cuales tienen mayores rendimientos.

**FIGURA 2.9A**

Histograma del porcentaje del rendimiento 2003 (Panel A—Fondos de crecimiento y Panel B—Fondos de valor).

**FIGURA 2.9B**

### Polígono

Es difícil y confuso realizar múltiples histogramas en la misma gráfica cuando comparamos dos o más conjuntos de datos. Al sobreponer las barras verticales de un histograma en otro se dificulta la interpretación. Cuando hay dos o más grupos, es conveniente utilizar un polígono de porcentaje.

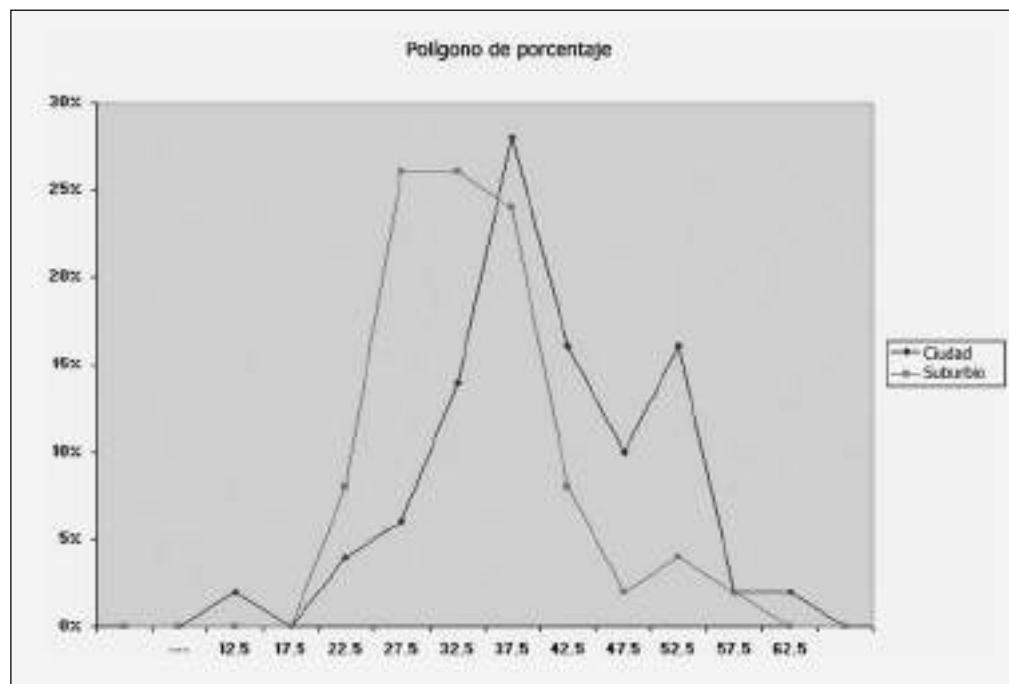
### POLÍGONO DE PORCENTAJE

El **polígono de porcentaje** se crea al hacer que el punto medio de cada clase represente los datos de esa clase y después se conecta la secuencia de puntos medios con sus respectivos porcentajes de clase.

La figura 2.10 muestra los polígonos de porcentaje para el costo de los alimentos en los restaurantes de la ciudad y de los suburbios. El polígono para los restaurantes de los suburbios está concentrado a la izquierda (correspondiente al costo menor) del polígono para los restaurantes de la ciudad. Los porcentajes de los costos más altos para los restaurantes de los suburbios corresponden a los puntos medios de clase de \$27.50 y \$32.50, mientras que los porcentajes de costo mayores para los restaurantes de la ciudad corresponden a los puntos medios de clase de \$37.50.

**FIGURA 2.10**

Polígonos de porcentaje del costo de las comidas para los restaurantes de la ciudad y de los suburbios.



Los polígonos de la figura 2.10 tienen puntos cuyos valores en el eje *X* representan el punto medio del intervalo de clase. Por ejemplo, observe los puntos trazados en el eje *X* en 22.5 (\$22.50). El punto para los restaurantes de los suburbios (el más alto) representa el hecho de que el 8% de estos restaurantes tienen costos por comida que van de los \$20 a los \$25. El punto para los restaurantes de la ciudad (el más bajo) representa el hecho de que el 4% de estos restaurantes tienen costos de comida entre \$20 y \$25.

Cuando elabore polígonos o histogramas, el eje vertical (*Y*) debe mostrar el verdadero cero u “origen”, para no distorsionar el carácter de los datos. El eje horizontal (*X*) no necesita especificar el punto cero para la variable de interés, aunque el rango de la variable debe constituir la mayor porción del eje.

### EJEMPLO 2.10

#### POLÍGONO DE PORCENTAJES PARA EL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

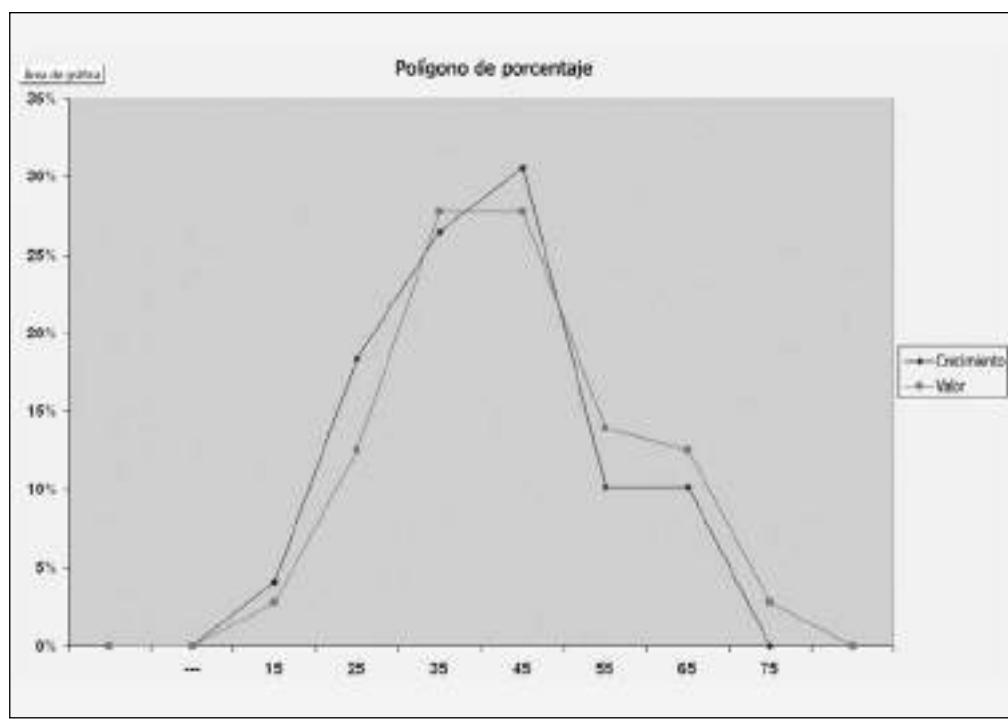
En el escenario de “Uso de la estadística”, se le pide que compare el rendimiento en 2003 de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Construya polígonos de porcentaje para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

La figura 2.11 muestra que la distribución de los fondos de crecimiento tiene un rendimiento anual menor en comparación con los fondos de valor, los cuales tienen mayores rendimientos.

**FIGURA 2.11**

Polígonos de porcentaje para el rendimiento en 2003.



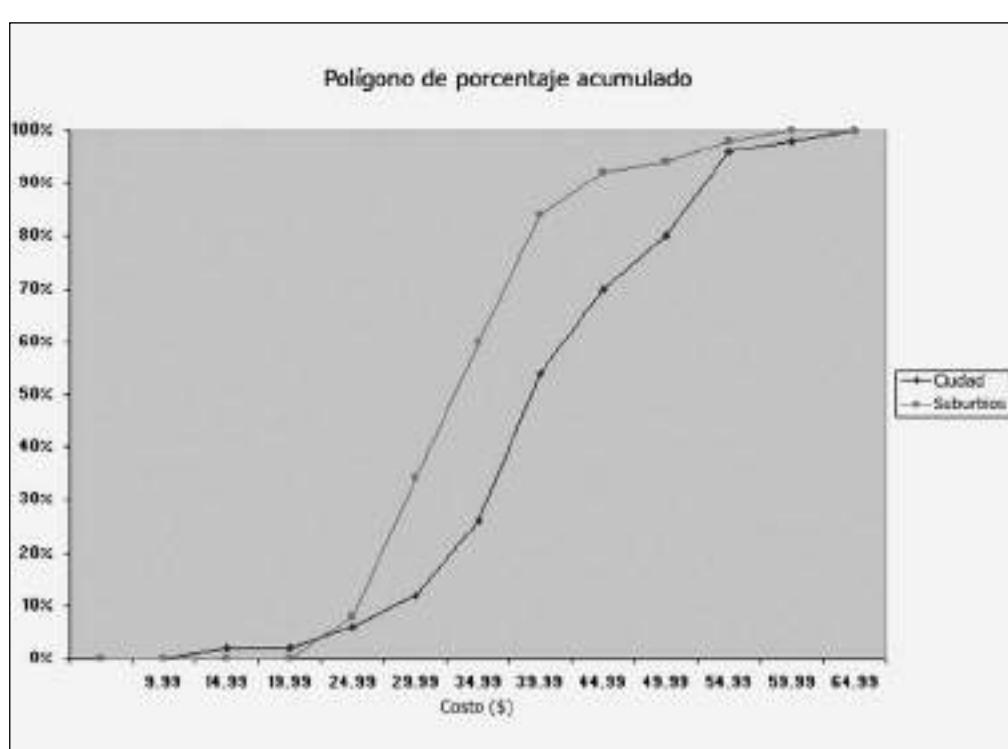
### Polígono de porcentaje acumulado (ojiva)

El **polígono de porcentaje acumulado**, u **ojiva**, muestra la variable de interés a lo largo del eje X, y los porcentajes acumulados a lo largo del eje Y.

La figura 2.12 ilustra los polígonos de porcentaje acumulado de Excel del costo de las comidas en los restaurantes de la ciudad y de los suburbios. La mayor parte de la curva correspondiente a los restaurantes de la ciudad está localizada a la derecha de la curva correspondiente a los restaurantes

**FIGURA 2.12**

Polígonos de porcentaje acumulado del costo de las comidas en los restaurantes de la ciudad y de los suburbios.



de los suburbios. Esto indica que los restaurantes de la ciudad tienen menos comidas que cuestan por debajo de un valor en particular. Por ejemplo, el 12% de las comidas de los restaurantes de la ciudad cuestan menos de \$30 en comparación con el 34% de las comidas de los restaurantes de los suburbios.

### EJEMPLO 2.11

#### POLÍGONOS DE PORCENTAJE ACUMULADO DE LOS RENDIMIENTOS EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

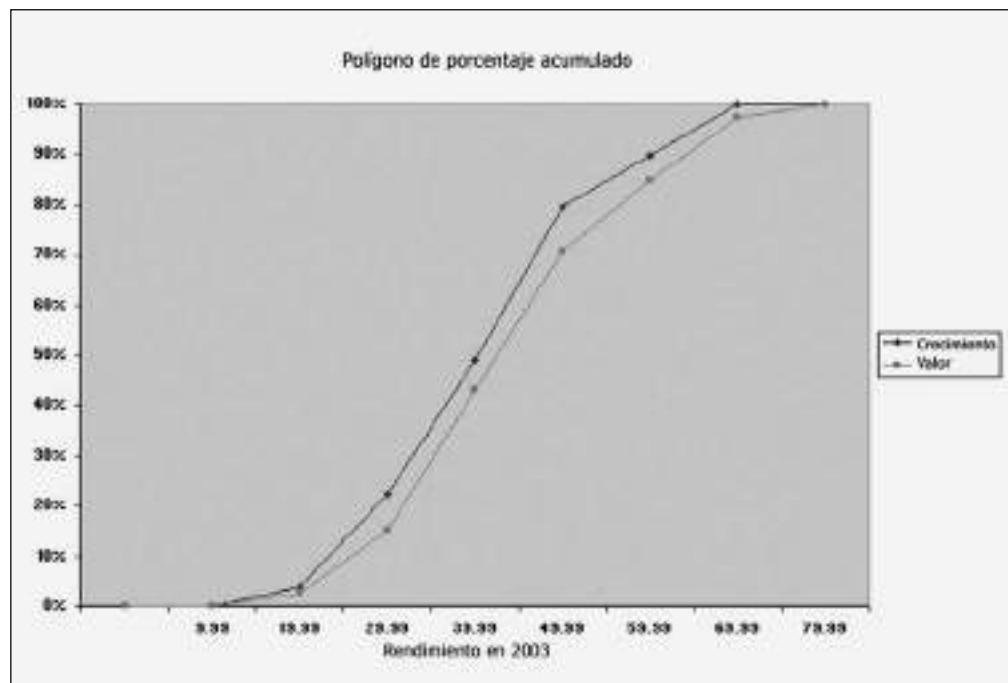
En el escenario “Uso de la estadística”, se le pide que compare el rendimiento de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Elabore polígonos de porcentaje acumulado para los fondos de crecimiento y para los fondos de valor.

#### SOLUCIÓN

La figura 2.13 ilustra los polígonos de porcentaje acumulado en Excel del porcentaje de rendimiento en 2003 de los fondos de crecimiento y de valor. La curva para los fondos de valor se localiza ligeramente a la derecha de la curva para los fondos de crecimiento. Esto indica que los fondos de valor tienen menos rendimientos por debajo de un valor específico. Por ejemplo, el 70.83% de los fondos de valor tienen rendimientos menores de 50 en comparación con el 79.59% de los fondos de crecimiento.

### FIGURA 2.13

Polígonos de porcentaje acumulado para el porcentaje de rendimiento en 2003.



## PROBLEMAS PARA LA SECCIÓN 2.3

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 2.20** Los valores para un conjunto de datos varían de 11.6 a 97.8.
- Si estos valores se agrupan en clases, indique los límites de clase.
  - ¿Qué amplitud de intervalo de clase eligió?
  - ¿Cuáles son los puntos medios de cada clase?

ASISTENCIA  
de PH Grade

- 2.21** Al realizar una ojiva (es decir, un polígono de porcentaje acumulado) relacionado con las puntuaciones del GMAT (siglas para Graduate Management Admission Test) de una muestra de 50 solicitantes para un programa de maestría en administración, los datos previos indicaron que ninguno de los solicitantes obtuvo puntuaciones por debajo de 450. La distribución de la frecuencia se formó eligiendo intervalos de clase 450 a 499, 500 a 549 y así sucesivamente.

mente, hasta que la última clase fue 700 a 749. Si dos solicitantes obtuvieron puntuaciones en el intervalo 450 a 499, y 16 solicitantes obtuvieron puntuaciones en el intervalo 500 a 549:

- ¿Qué porcentaje de solicitantes calificó por debajo de 500?
- ¿Qué porcentaje de solicitantes calificó entre 500 y 549?
- ¿Qué porcentaje de solicitantes calificó por debajo de 550?
- ¿Qué porcentaje de solicitantes calificó por debajo de 750?

### Aplicación de conceptos

Puede resolver manualmente los problemas 2.22 a 2.27 o usando Excel, Minitab o SPSS.

**ASISTENCIA** **2.22** Los datos mostrados a continuación representan el costo de la energía eléctrica durante julio de 2004 para una muestra aleatoria de 50 departamentos de una habitación en una gran ciudad. **UTILITY**

#### Datos brutos de los cargos de utilidades (\$)

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

- Forme una distribución de frecuencia y una distribución de porcentajes que tenga intervalos de clase con los límites superiores de clase \$99, \$119, y así sucesivamente.
- Trace un histograma y un polígono de porcentaje.
- Elabore una distribución de porcentaje acumulado y trace una ojiva (polígono de porcentaje acumulado).
- ¿Alrededor de cuál cantidad parece concentrarse el costo mensual de la energía eléctrica?



**2.23** Una de las operaciones que realiza un molino consiste en cortar piezas de acero en partes que posteriormente serán usadas como marco para los asientos delanteros de un automóvil. El acero se corta con una sierra con punta de diamante y se requiere que las partes resultantes midan  $\pm 0.005$  pulgadas de longitud, según las especificaciones de la empresa automovilística. La siguiente tabla proviene de una muestra de 100 partes de acero. La medida reportada es la diferencia en pulgadas entre la longitud real de la parte de acero, medida con un dispositivo láser y la longitud especificada de la parte de acero. Por ejemplo, el primer valor,  $-0.002$ , representa una parte de acero que es 0.002 pulgadas más corta que la longitud especificada. **STEEL**

-0.002	0.002	0.0005	-0.0015	-0.001
0.0005	0.001	0.001	-0.0005	-0.001
0.0025	0.001	0.0005	-0.0015	0.0005
0.001	0.001	0.001	-0.0005	-0.0025
0.002	-0.002	0.0025	-0.0005	0.0025
0.001	-0.003	0.001	-0.001	0.002
0.005	-0.0015	0	-0.0015	0.0025

-0.002	-0.0005	-0.0025	0.0025	-0.002
0	0	-0.001	0.001	0
0.001	-0.0025	0.0035	0.0005	-0.0005
-0.0025	-0.003	0	0	-0.001
-0.003	-0.001	-0.003	0.002	0
0.001	0.002	-0.002	-0.0005	-0.002
-0.0005	-0.001	-0.001	0.0005	0
0	0	-0.0015	0.0005	0
-0.003	0.003	-0.0015	0	0.002
-0.001	0.0015	-0.002	-0.0005	-0.003
0.0005	0	0.001	0.002	-0.0005
0.0025	0	-0.0025	0.001	-0.002
-0.0025	-0.0025	-0.0005	-0.0015	-0.002

- Realice una distribución de frecuencias y una distribución de porcentaje.
- Trace un histograma y un polígono de porcentaje.
- Trace un polígono de porcentaje acumulado.
- ¿El molino realiza un buen trabajo, de acuerdo con los requerimientos de la empresa automovilística? Explique su respuesta.

**2.24** Una compañía productora fabrica bastidores de acero para equipos eléctricos. El componente principal de los bastidores es un canalón de acero enrollado de calibre 14. Éste se produce con una prensa cuyo poder de golpeo progresivo es de 250 toneladas, con una operación de limpieza hacia abajo que pone dos formas de 90 grados en el acero aplanado para hacer el canalón. La distancia de un lado de la forma al otro es importante por la resistencia a exteriores. La compañía requiere que la amplitud del canalón esté entre 8.31 y 8.61 pulgadas. Los siguientes datos son las longitudes de los canalones en pulgadas para una muestra de  $n = 49$ . **TROUGH**

8.312	8.343	8.317	8.383	8.348	8.410	8.351	8.373
8.481	8.422	8.476	8.382	8.484	8.403	8.414	8.419
8.385	8.465	8.498	8.447	8.436	8.413	8.489	8.414
8.481	8.415	8.479	8.429	8.458	8.462	8.460	8.444
8.429	8.460	8.412	8.420	8.410	8.405	8.323	8.420
8.396	8.447	8.405	8.439	8.411	8.427	8.420	8.498
8.409							

- Realice una distribución de frecuencia y una distribución de porcentajes.
- Trace un histograma y un polígono de porcentajes.
- Trace un polígono de porcentajes acumulados.
- ¿Qué puede concluir acerca del número de canalones que satisfarán los requerimientos de la compañía, es decir, que tengan una longitud entre 8.31 y 8.61 pulgadas?

**2.25** La compañía productora del problema 2.24 también fabrica aislantes eléctricos. Si éstos se descomponen cuando están en uso, es probable que ocurra un corto circuito. Para probarlos, se efectúa una prueba destructiva en laboratorios de alta potencia, que determinarán cuánta *fuerza* se requiere para descomponer los aislantes. La fuerza se mide observando cuántas libras deben aplicarse al aislante antes de que se descomponga. La fuerza de 30 aislantes probados se muestra a continuación. **FORCE**

1,870 1,728 1,656 1,610 1,634 1,784 1,522 1,696  
 1,592 1,662 1,866 1,764 1,734 1,662 1,734 1,774  
 1,550 1,756 1,762 1,866 1,820 1,744 1,788 1,688  
 1,810 1,752 1,680 1,810 1,652 1,736

- Construya una distribución de frecuencias y una distribución de porcentajes.
- Trace un histograma y un polígono de porcentajes.
- Trace un polígono de porcentaje acumulado.
- ¿Qué puede concluir respecto de la fuerza de los aislantes, si la compañía requiere una medida de fuerza de por lo menos 1,500 libras antes de descomponerse?

**2.26** Los arreglos ordenados de la tabla corresponden a la vida útil (en horas) de una muestra de 40 bulbos de 100 watts producidos por el fabricante A y la muestra B a 40 bulbos de 100 watts producidos por otro fabricante. **BULBS**

Fabricante A					Fabricante B				
684	697	720	773	821	819	836	888	897	903
831	835	848	852	852	907	912	918	942	943
859	860	868	870	876	952	959	962	986	992
893	899	905	909	911	994	1,004	1,005	1,007	1,015
922	924	926	926	938	1,016	1,018	1,020	1,022	1,034
939	943	946	954	971	1,038	1,072	1,077	1,077	1,082
972	977	984	1,005	1,014	1,096	1,100	1,113	1,113	1,116
1,016	1,041	1,052	1,080	1,093	1,153	1,154	1,174	1,188	1,230

- Realice una distribución de frecuencias y una distribución de porcentajes para cada fabricante usando el siguiente intervalo de clase para cada distribución:

- Fabricante A: 650 pero menos de 750, 750 pero menos de 850, y así sucesivamente.
- Fabricante B: 750 pero menos de 850, 850 pero menos de 950, y así sucesivamente.

- Trace los histogramas de porcentaje en gráficas separadas y trace los polígonos de porcentaje en una gráfica.

- Elabore las distribuciones de porcentaje acumulado y trace las ojivas en una gráfica.
- ¿Qué fabricante produce los bulbos con mayor vida: el fabricante A o el fabricante B? Explique su respuesta.

**2.27** Los siguientes datos representan la cantidad de bebida gaseosa en una muestra de 50 botellas de 2 litros. **DRINK**

2.109 2.086 2.066 2.075 2.065 2.057 2.052 2.044  
 2.036 2.038 2.031 2.029 2.025 2.029 2.023 2.020  
 2.015 2.014 2.013 2.014 2.012 2.012 2.012 2.010  
 2.005 2.003 1.999 1.996 1.997 1.992 1.994 1.986  
 1.984 1.981 1.973 1.975 1.971 1.969 1.966 1.967  
 1.963 1.957 1.951 1.951 1.947 1.941 1.941 1.938

1.908 1.894

- Construya una distribución de frecuencias y una distribución de porcentajes.

- Trace un histograma y un polígono de porcentajes.

- Realice una distribución de porcentaje acumulado y trace un polígono de porcentaje acumulado.

- Con base en los resultados de a) a c), ¿la cantidad de bebida gaseosa con que se llenan las botellas se concentra alrededor de valores específicos?

## 2.4 TABLAS Y GRÁFICAS DE DATOS BIVARIADOS

En negocios es común el estudio de patrones que pueden existir entre dos o más variables categóricas.

### Tabla de contingencia

Una **tabla de clasificación** (o **contingencia**) **cruzada** presenta los resultados de dos variables categóricas. Las respuestas en conjunto se clasifican de tal manera que las categorías de una variable se localizan en las filas, y las categorías de la otra variable se localizan en las columnas. Los valores localizados en las intersecciones de las filas y las columnas se llaman **celdas**. La tabla se construye dependiendo del tipo de contingencia, las celdas para cada combinación de fila-columna contienen la frecuencia, el porcentaje del total global, el porcentaje del total de las filas o el porcentaje total de las columnas.

Suponga que en el escenario de “Uso de la estadística” se quiere examinar si hay o no un patrón o relación entre el nivel de riesgo y el objetivo del fondo de inversión (crecimiento contra valor). La tabla 2.14 resume esta información para los 121 fondos de inversión.

**TABLA 2.14**

Tabla de contingencia que muestra el fondo objetivo y el fondo de riesgo.

OBJETIVO	NIVEL DE RIESGO			Total
	Alto	Promedio	Bajo	
Crecimiento	14	23	12	49
Valor	3	23	46	72
Total	17	46	58	121

Se elaboró esta tabla de contingencia etiquetando las respuestas en conjunto para cada uno de los 121 fondos de inversión con respecto al objetivo y al riesgo en una de las seis posibles celdas en la tabla. Así, el primer fondo en la lista (AFBA Five Star USA Global Institutional) está clasificado como fondo de crecimiento con riesgo promedio. Por tanto, registre la respuesta conjunta dentro de la celda que forma la intersección de la primera fila y la segunda columna. Las 120 respuestas conjuntas restantes se registran de forma similar. Cada celda contiene la frecuencia para la combinación fila-columna.

Para explorar cualquier posible patrón o relación entre fondos objetivos y de riesgo, es conveniente realizar tablas de contingencia basadas en porcentajes. Primero convierta en porcentajes estos resultados con base en los siguientes tres totales:

1. El total global (es decir, los 121 fondos de inversión).
2. El total de las filas (es decir, 49 fondos de crecimiento y 72 fondos de valores).
3. El total de las columnas (es decir, los tres niveles de riesgo).

Las tablas 2.15, 2.16 y 2.17 resumen estos porcentajes.

**TABLA 2.15**

Tabla de contingencia que muestra el fondo objetivo y el fondo de riesgo con base en el porcentaje del total global.

OBJETIVO	NIVEL DE RIESGO			Total
	Alto	Promedio	Bajo	
Crecimiento	11.57	19.01	9.92	40.50
Valor	2.48	19.01	38.02	59.50
Total	14.05	38.02	47.93	100.00

**TABLA 2.16**

Tabla de contingencia que muestra el fondo objetivo y el fondo de riesgo con base en el porcentaje del total de las filas.

OBJETIVO	NIVEL DE RIESGO			Total
	Alto	Promedio	Bajo	
Crecimiento	28.57	46.94	24.49	100.00
Valor	4.17	31.94	63.89	100.00
Total	14.05	38.02	47.93	100.00

**TABLA 2.17**

Tabla de contingencia que muestra el fondo objetivo y el fondo de riesgo con base en el porcentaje del total de las columnas.

OBJETIVO	NIVEL DE RIESGO			Total
	Alto	Promedio	Bajo	
Crecimiento	82.35	50.00	20.69	40.50
Valor	17.65	50.00	79.31	59.50
Total	100.00	100.00	100.00	100.00

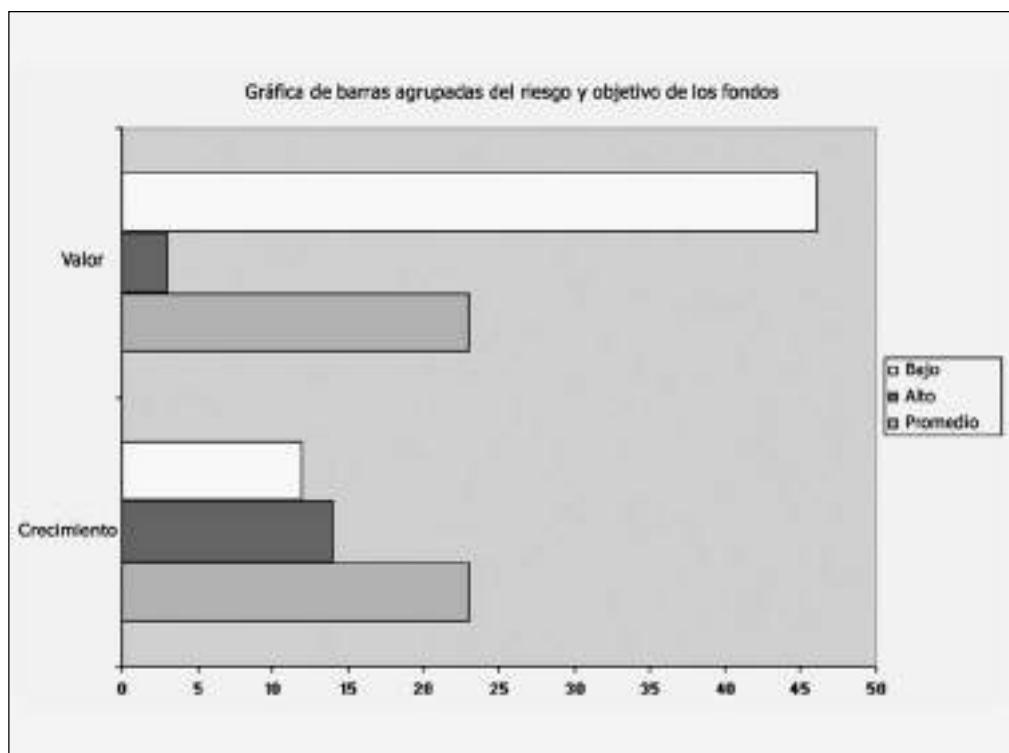
La tabla 2.15 muestra que el 14.05% de los fondos de inversión de la muestra son de alto riesgo, el 40.5% son fondos de crecimiento y el 11.57% son fondos de crecimiento de alto riesgo. La tabla 2.16 muestra que el 28.57% de los fondos de crecimiento son de alto riesgo y el 24.49% son de bajo riesgo. La tabla 2.17 muestra que el 82.35% de los fondos de alto riesgo y sólo el 20.69% de los fondos de bajo riesgo son fondos de crecimiento. Las tablas revelan que los fondos de crecimiento tienen mayor probabilidad de ser de alto riesgo, mientras que los fondos de valor tienen mayor probabilidad de ser de bajo riesgo.

## Gráfica de barras agrupadas

Una forma útil de mostrar los resultados de datos de clasificación cruzada es realizar una **gráfica de barras agrupadas**. La figura 2.14, que utiliza datos de la tabla 2.14, es una gráfica de barras agrupadas de Excel que compara los tres niveles de riesgo de los fondos, con base en su objetivo. Al examinar la figura 2.14, se revelan resultados congruentes con aquellos de las tablas 2.15, 2.16 y 2.17. Los fondos de crecimiento tienen mayor probabilidad de ser de alto riesgo, mientras que los fondos de valor tienen mayor probabilidad de ser de bajo riesgo.

**FIGURA 2.14**

Gráfica de barras agrupadas de Excel para el objetivo y riesgo de los fondos.



## PROBLEMAS PARA LA SECCIÓN 2.4

### Aprendizaje básico

**ASISTENCIA de PH Grade** **2.28** Los siguientes datos representan las respuestas a dos preguntas de una encuesta a 40 alumnos que estudian la especialidad en negocios: ¿Cuál es tu género? (Masculino = M; Femenino = F) y ¿Cuál es tu especialidad? Contaduría = A; Sistemas computacionales de la información = C; Marketing = M):

Género: M M M F M F F M F M F M M M M F F M F F

Especialidad: A C C M A C A A C C A A A M C M A A A C

Género: M M M M F M F F M M F M M M M F M F M M

Especialidad: C C A A M M C A A A C C A A A A C C A C

- Registre los datos en una tabla de contingencia, donde las dos filas representen la categoría de género y las tres columnas representen la categoría de la especialidad académica.
- Elabore tablas de contingencia a partir de los porcentajes de las respuestas de los 40 estudiantes, con base en los porcentajes de las filas y en los porcentajes de las columnas.

- Usando los resultados del inciso a), construya una gráfica de barras agrupadas de género basada en la especialidad del estudiante.

**ASISTENCIA de PH Grade** **2.29** A partir de la siguiente tabla de contingencia, elabore una gráfica de barras agrupadas, comparando A y B para cada una de las categorías de tres columnas en el eje vertical.

	1	2	3	Total
A	20	40	40	100
B	80	80	40	200

### Aplicando los conceptos

**ASISTENCIA de PH Grade** **✓ AUTO Examen** **2.30** Los resultados de un estudio realizado como parte de un esfuerzo por mejorar la producción en una fábrica de semiconductores presentan datos de defectos en una muestra de 450 placas de silicio. La siguiente tabla presenta un resumen de las respues-

tas a dos preguntas: ¿Se encontró una partícula en el troquel que produjo la placa de silicio? Y ¿La placa resultó buena o mala?

CALIDAD DE LA PLACA	CONDICIÓN DEL TROQUEL		
	Sin partículas	Partículas	Totales
Buena	320	14	334
Mala	80	36	116
Total	400	50	450

Fuente: S.W. Hall, *Analysis of Defectivity of Semiconductor Wafers by Contingency Table*, Proceedings Institute of Environmental Sciences. Vol.1 (1994), 177-183.

- Construya tablas de contingencia basadas en el total de porcentajes, porcentajes de fila y porcentajes de columna.
- Elabore una gráfica de barras agrupadas de la calidad de las placas de silicio basada en la condición del troquel.
- ¿A qué conclusiones llega a partir de esos análisis?

**ASISTENCIA de PH Grade** **2.31** En un gran hospital cada día se realizan varios cientos de pruebas de laboratorio. La tasa de pruebas realizadas de forma impropia (y que deben practicarse de nuevo) al parecer es constante y cercana al 4%. En un esfuerzo por llegar a la raíz del problema (pruebas que requieren volver a hacerse), el director del laboratorio decide guardar los registros de una semana. Las pruebas de laboratorio se dividieron entre el turno de empleados que realizan las pruebas de laboratorio. Los resultados son los siguientes:

PRUEBAS DE LABORATORIO REALIZADAS	TURNO		
	Día	Tarde	Total
Insatisfactorias	16	24	40
Satisfactorias	654	306	960
Total	670	330	1,000

- Elabore tablas de contingencia basadas en los porcentajes totales, los porcentajes de filas y los porcentajes de columnas.
- ¿Qué tipo de porcentaje (de fila, de columna o total) considera que es el más informativo para estos datos? Explique su respuesta.
- ¿A qué conclusiones llegaría el director del laboratorio, respecto al patrón de pruebas de laboratorio insatisfactorias?

**2.32** Se seleccionó una muestra de 500 compradores en una amplia área metropolitana, para determinar información variada con relación al comportamiento del consumidor. Entre las preguntas que se hicieron estaba “¿Disfruta comprando ropa?” Los resultados se resumen en la siguiente tabla de contingencia:

DISFRUTA COMPRANDO ROPA	GÉNERO		
	Masculino	Femenino	Total
Sí	136	224	360
No	104	36	140
Total	240	260	500

- Elabore tablas de contingencia basadas en los porcentajes totales, los porcentajes de fila y los porcentajes de columna.
- Construya una gráfica de barras agrupadas en Disfruta comprando ropa basada en el género.

- ¿A qué conclusiones llega a partir de estos análisis?

**2.33** Las ventas al detalle en Estados Unidos para abril de 2002, fueron ligeramente superiores a las de abril de 2001. Todas las tiendas de descuento, como Wal Mart, Costco, Target y Dollar General, incrementaron sus ventas en un 9% o más. Sin embargo, las ventas al detalle en la industria de la ropa fueron mixtas. La siguiente tabla presenta el total de las ventas al detalle en millones de dólares para las compañías de ropa líderes durante abril de 2001 y abril de 2002.

COMPAÑÍA DE ROPA	VENTAS TOTALES EN MILLONES DE DÓLARES	
	Abril 01	Abril 02
Gap	1,159.0	962.0
TJX	781.7	899.0
Limited	596.5	620.4
Kohl's	544.9	678.9
Nordstrom	402.6	418.3
Talbots	139.9	130.1
AnnTaylor	114.2	124.8

Fuente: Ann Zimmerman, “Retail Sales Grow Modestly”, The Wall Street Journal, 10 de mayo, 2002, B4.

- Realice una tabla de los porcentajes de columnas.
- Elabore una gráfica de barras agrupadas para visualizar lo relevante de la información recabada en el inciso a).
- Analice los cambios de las compras al detalle para la industria de la ropa entre abril de 2001 y abril de 2002.

**2.34** Con el fin de estimular las ventas de 2003, los fabricantes de autos ofrecieron grandes incentivos, en forma de rebajas en efectivo para los compradores de autos nuevos. Por ejemplo, los compradores de autos de marca Lincoln recibieron un promedio de rebajas de \$4,086. A pesar de tales rebajas, los fabricantes estadounidenses de autos perdieron una parte del mercado global en favor de la competencia internacional.

MARCA	REBAJAS DE EFECTIVO (EN DÓLARES)	
	2001	2003
Buick	1,939	3,655
Chevrolet	1,654	3,231
Chrysler	1,835	2,832
Ford	1,334	2,752
Lincoln	2,449	4,086

Fuente: K. Lundgaard y S. Freeman, “Detroit's Challenge: Weaning Buyers from Years of Deals”, The Wall Street Journal, 6 de enero, 2004, A1.

- Elabore una gráfica de barras agrupadas para las cinco marcas.
- Analice los cambios en el tamaño de las rebajas en efectivo de 2001 a 2003.

**2.35** La venta de autos en Estados Unidos se incrementó un 3.3% en enero de 2004 en comparación con enero de 2003. Los fabricantes japoneses de automóviles experimentaron un incremento mucho mayor. La siguiente tabla contiene las ventas de autos y camiones ligeros de algunos de los grandes fabricantes durante enero de 2003 y 2004.

VENTAS DE AUTOS NUEVOS Y CAMIONES LIGEROS		
FABRICANTE	2003	2004
Nissan	55,213	72,164
Honda	89,993	90,173
Toyota	119,376	143,729
Chrysler	144,826	162,205
Ford	242,068	229,238
GM	291,254	296,788

Fuente: S. Freeman y J. B. White, "U. S. Car Sales Rose 3.3% in January", The Wall Street Journal, 4 de febrero, 2004, A2.

- Elabore una gráfica de barras agrupadas para las seis marcas.
- Discuta los cambios en las ventas de autos nuevos y camiones ligeros en enero de 2004 comparadas con enero de 2003.

## 2.5 DIAGRAMAS DE DISPERSIÓN Y SERIES DE TIEMPO

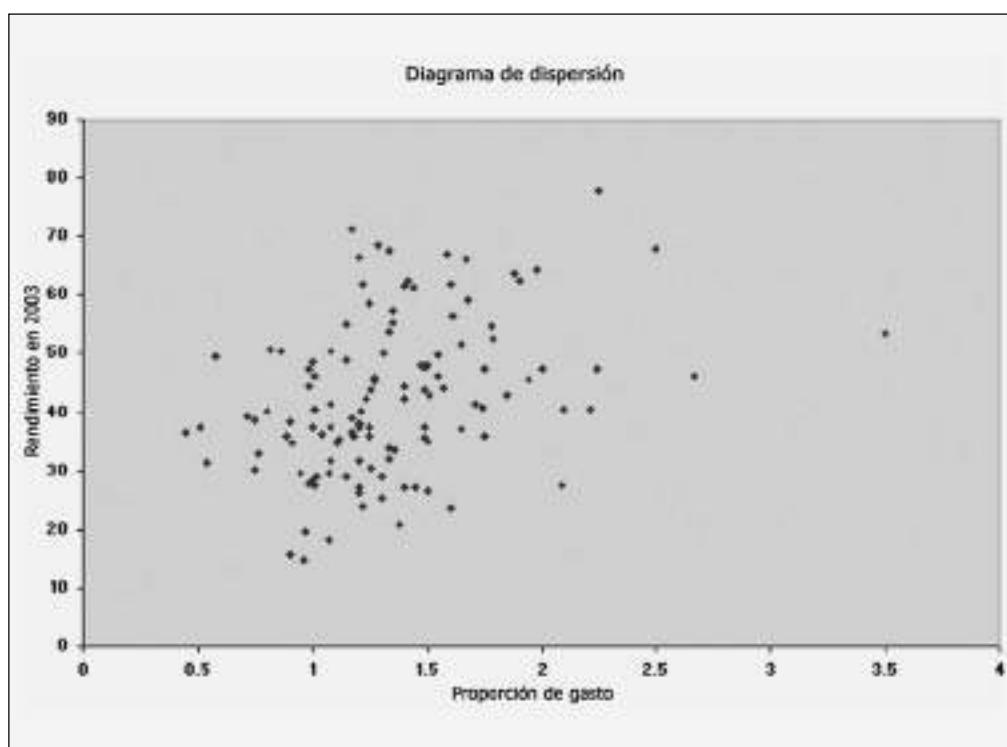
### Diagrama de dispersión

Al analizar una sola variable numérica, tal como el costo de la comida en un restaurante o el rendimiento en 2003, se usa un histograma, un polígono o un polígono de porcentaje acumulado como los desarrollados en la sección 2.3. Utilice un **diagrama de dispersión** para examinar las posibles relaciones entre dos variables numéricas. Coloque una variable en el eje horizontal *X* y la otra variable en el eje vertical *Y*. Por ejemplo, un analista de mercado podría estudiar la efectividad de la publicidad si compara los volúmenes de ventas semanales y el gasto publicitario semanal. O un director de recursos humanos interesado en la estructura del salario de una empresa podría comparar los años de experiencia de los empleados y su salario actual.

Para mostrar el diagrama de dispersión, habría que estudiar la relación entre la proporción del gasto y el rendimiento en 2003. Para cada fondo de inversión, se traza la proporción de gasto en el eje horizontal *X*, y el rendimiento en 2003 en el eje vertical *Y*. La figura 2.15 representa la salida de Excel para estas dos variables.

**FIGURA 2.15**

Diagrama de dispersión de Excel para la proporción de gasto y el rendimiento en 2003.



Aunque hay una gran variación en la proporción del gasto y el rendimiento en 2003 de los fondos de inversión, parece haber una relación creciente (positiva) entre la proporción del gasto y el

rendimiento en 2003. En otras palabras, los fondos que tienen una baja proporción de gasto tienen un bajo rendimiento en 2003. Quizá otros pares de variables tengan relaciones decrecientes (negativas) en las que una variable decrece a medida que la otra se incrementa. El diagrama de dispersión se volverá a estudiar en el capítulo 13, cuando desarrollemos el análisis de regresión.

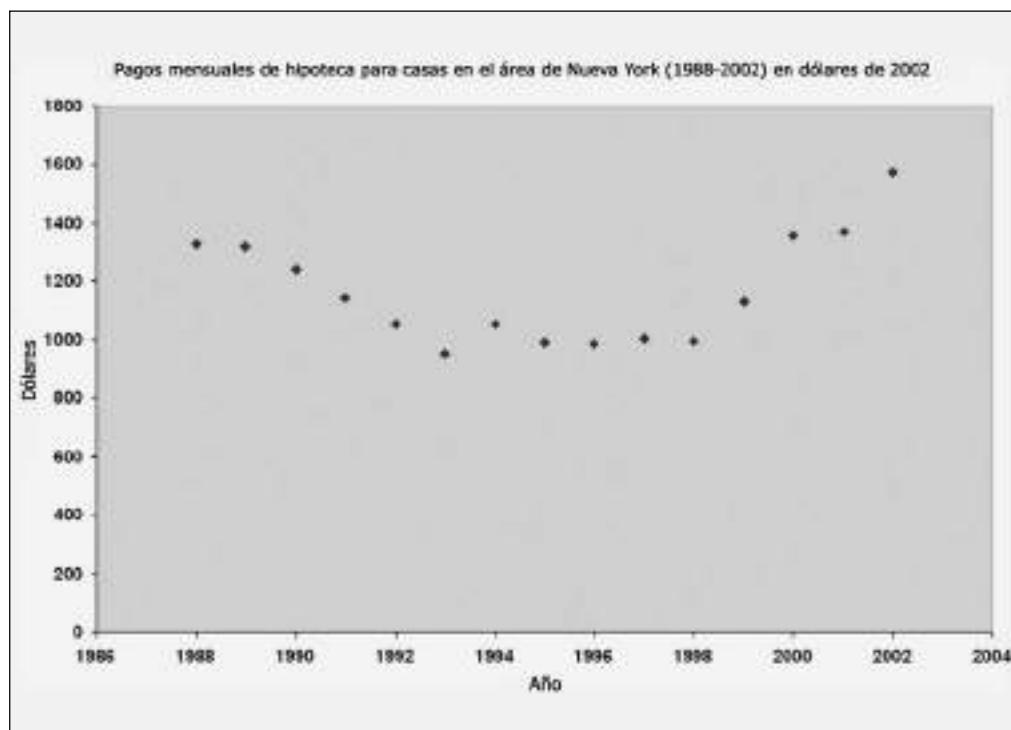
## Series de tiempo

El **diagrama de series de tiempo** se usa para estudiar patrones en las variables a través del tiempo. Cada valor se traza como un punto de dos dimensiones. Un diagrama de series de tiempo muestra el periodo de tiempo en el eje horizontal *X* y la variable de interés en el eje vertical *Y*.

La figura 2.16 es un diagrama de series de tiempo para el pago mensual de hipoteca (en dólares de 2002) de 1988 a 2002. **HOUSESNY**

**FIGURA 2.16**

Diagrama de series de tiempo de Excel del pago mensual de hipoteca en dólares de 2002 (periodo de 1988 a 2002).



Los pagos mensuales de hipoteca (considerados en dólares de 2002) bajaron al final de los 80 y principios de los 90, sólo para nivelarse. Comenzaron a aumentar de nuevo a partir de 1999.

## PROBLEMAS PARA LA SECCIÓN 2.5

### Aprendizaje básico

**ASISTENCIA de PH Grade** **2.36** El siguiente es un conjunto de datos tomados de una muestra de  $n = 11$  artículos.

<i>X</i>	7	5	8	3	6	10	12	4	9	15	18
<i>Y</i>	21	15	24	9	18	30	36	12	27	45	54

- Trace un diagrama de dispersión.
- ¿Existe una relación entre *X* y *Y*? Explique su respuesta.

ASISTENCIA  
de PH Grade

**2.37** La siguiente es una serie de ventas anuales reales (en millones de dólares constantes de 1995) en un periodo de 11 años (1992 a 2002).

<i>Año</i>	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
<i>Ventas</i>	13.0	17.0	19.0	20.0	20.5	20.5	20.5	20.0	19.0	17.0	13.0

- Realice un diagrama de series de tiempo.
- ¿Parece haber algún cambio en las ventas anuales reales a través del tiempo? Explique su respuesta.

## Aplicación de conceptos

Puede resolver manualmente los problemas 2.38 a 2.45 o usando Excel, Minitab o SPSS.

**2.38** Los siguientes datos representan el precio aproximado (en dólares) de las ventas al detalle y el costo de la energía por año (en dólares) de 15 refrigeradores. **REFRIGERATOR**

Modelo	Precio	Costo de energía
Maytag MTB1956GE	825	36
Kenmore 7118	750	43
Maytag MTB2156GE	850	39
Kenmore Elite	1000	38
Amana ART2107B	800	38
GE GTS18KCM	600	40
Kenmore 7198	750	35
Frigidaire Gallery GLHT216TA	680	38
Kenmore 7285	680	40
Whirlpool Gold GR9SHKXK	940	37
Frigidaire Gallery GLRT216TA	680	40
GE GTS22KCM	650	44
Whirlpool ETF1TTXK	800	43
Whirlpool Gold GR2SHXK	1050	40
Frigidaire FRT18P5A	510	40

Fuente: "Refrigerators", Copyright 2002 by Consumers Union of U.S., Inc., Yonkers, NY 10703-1057, a nonprofit organization. Adaptado con permiso de Consumer Reports, 26 de agosto, 2002, para propósitos educativos exclusivamente. No se permite su uso comercial o reproducción. [www.ConsumerReports.org](http://www.ConsumerReports.org)

- a. Elabore un diagrama de dispersión con el costo de energía en el eje X y el precio en el eje Y.
- b. ¿Parece haber relación entre el precio y el costo de la energía? De ser así, ¿la relación es positiva o negativa?
- c. ¿Esperaría que los refrigeradores con un precio más alto tengan mayor eficiencia de energía? ¿Los datos apoyan esto?

**2.39** Los siguientes datos **SECURITY** representan la proporción del volumen de ventas de las pantallas de preabordaje en los aeropuertos en 1998 y 1999 y las violaciones de seguridad detectadas por millones de pasajeros.

Ciudad	Volumen de ventas	Violaciones
St. Louis	416	11.9
Atlanta	375	7.3
Houston	237	10.6
Boston	207	22.9
Chicago	200	6.5
Denver	193	15.2
Dallas	156	18.2
Baltimore	155	21.7
Seattle/Tacoma	140	31.5
San Francisco	110	20.7

Ciudad	Volumen de ventas	Violaciones
Orlando	100	9.9
Washington-Dulles	90	14.8
Los Ángeles	88	25.1
Detroit	79	13.5
San Juan	70	10.3
Miami	64	13.1
Nueva York-JFK	53	30.1
Washington-Reagan	47	31.8
Honolulu	37	14.9

Fuente: Alan B. Krueger, "A Small Dose of Common Sense Would Help Congress Break the Gridlock over Airport Security", The New York Times, 15 de noviembre, 2001, C2.

- a. Elabore un diagrama de dispersión con la proporción de volumen de ventas de las pantallas de preabordaje en el eje X y las violaciones de seguridad detectadas en el eje Y.

- b. ¿A qué conclusiones llega acerca de la relación entre la proporción del volumen de ventas de las pantallas de preabordaje y las violaciones de seguridad detectadas?

**2.40** Los siguientes datos **CELLPHONE** representan el tiempo de llamada en horas en el modo digital y la capacidad de la batería en horas-miliamperes de los teléfonos celulares.

Tiempo de llamada	Capacidad de las baterías	Tiempo de llamada	Capacidad de las baterías
4.50	800	1.50	450
4.00	1500	2.25	900
3.00	1300	2.25	900
2.00	1550	3.25	900
2.75	900	2.25	700
1.75	875	2.25	800
1.75	750	2.50	800
2.25	1100	2.25	900
1.75	850	2.00	900

Fuente: "Service Shortcomings", Copyright 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, febrero de 2002, 25.

- a. Realice un diagrama de dispersión con la capacidad de las baterías en el eje X y el modo digital del tiempo de llamada en el eje Y.

- b. ¿A qué conclusiones llega acerca de la relación entre la capacidad de la batería y el modo digital del tiempo de llamada?

- c. ¿Esperaría que los teléfonos celulares con mayor capacidad de batería tengan un mayor tiempo de llamada? ¿Apoyan esto los datos?

**2.41** Los siguientes datos **BATTERIES2** representan el precio y la corriente inicial que pueden generar las baterías de los automóviles.

Nombre	Precio (\$)	Ci
NAPA Legend Professional Line 7575	60	630
Exide Nascar Select 75-84N	80	630
DieHard Weatherhandler 30375 (South)	60	525
DieHard Weatherhandler 30075 (North)	60	650
EverStart 75-5	30	525
Duralast 75-D	50	650
Interstate Mega-Tron MT-75	80	650
EverStart 75-2	60	650
ACDelco Maintenance free 75A-72	80	650
Motorcraft Premier Silver Series BXT-75	80	700
DieHard Gold 33165 (South)	80	700
EverStart Extreme 65-2N (North)	60	850
ACDelco Maintenance Free 65-84	92	850
Exide 65-60	85	850
EverStart Extreme 65-2 (South)	60	675
DieHard Gold 33065 (North)	80	900
Duralast Gold 34DT-DGS (South)	70	800
Duralast Gold 34DT-DGN (North)	70	900
Interstate Mega-Tron Plus MTP-78DT	96	800
Optima Red Top 34/78-1050	140	750
ACDelco Professional 78DT-7YR	80	850
EverStart High Power DT-3	40	630
DieHard Weatherhandler 30034 (North)	60	540
DieHard Weatherhandler 30334 (South)	60	525

Fuente: "Leading the Charge", Copyright 2001 por Consumers Union of U.S. Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, octubre de 2001, 25.

- Elabore un diagrama de dispersión con la corriente inicial en el eje X y el precio en el eje Y.
- ¿A qué conclusiones llega respecto a la relación entre la corriente inicial y el precio?
- ¿Esperaría que las baterías con mayor corriente inicial tengan un mayor precio? ¿Los datos apoyan esto?

**2.42** El U.S. Bureau of Labor Statistics compila datos de una gran variedad de temas sobre la fuerza de trabajo. La siguiente tabla muestra la tasa de desempleo ajustada mensualmente por temporada para Estados Unidos, desde 1998 hasta 2003. **UERATE**

Mes	Tasa de desempleo en E.U. ajustada por temporada (en porcentaje)					
	1998	1999	2000	2001	2002	2003
Enero	4.7	4.3	4.0	4.2	5.6	5.9
Febrero	4.6	4.4	4.1	4.2	5.6	6.0
Marzo	4.7	4.2	4.0	4.3	5.7	5.9
Abril	4.3	4.3	4.0	4.5	5.9	6.1
Mayo	4.4	4.2	4.1	4.4	5.8	6.2
Junio	4.5	4.3	4.0	4.5	5.8	6.4
Julio	4.5	4.3	4.0	4.5	5.8	6.3
Agosto	4.5	4.2	4.1	4.9	5.8	6.2
Septiembre	4.5	4.2	3.9	4.9	5.7	6.2
Octubre	4.5	4.1	3.9	5.4	5.8	6.1
Noviembre	4.4	4.1	4.0	5.6	5.9	6.1
Diciembre	4.4	4.1	4.0	5.8	6.0	5.8

Fuente: U.S. Bureau of Labor Statistics.

- Construya un diagrama de series de tiempo para la tasa de desempleo de Estados Unidos.
- ¿Parece haber algún patrón?

**2.43** Los siguientes datos **DRINK** representan la cantidad de bebida gaseosa en una muestra consecutiva de 50 botellas de dos litros. Los resultados se enuncian horizontalmente en el orden en el que fueron llenados.

2.109 2.086 2.066 2.075 2.065 2.057 2.052 2.044 2.036 2.038  
2.031 2.029 2.025 2.029 2.023 2.020 2.015 2.014 2.013 2.014  
2.012 2.012 2.012 2.010 2.005 2.003 1.999 1.996 1.997 1.992  
1.994 1.986 1.984 1.981 1.973 1.975 1.971 1.969 1.966 1.967  
1.963 1.957 1.951 1.951 1.947 1.941 1.941 1.938 1.908 1.894

- Realice un diagrama de series de tiempo para la cantidad de bebida gaseosa en el eje Y y el número de botellas (de 1 a 50 consecutivamente) en el eje X.
- ¿Qué patrón, si lo hay, se presenta en los datos?
- Si tuviera que hacer una predicción de la cantidad de bebida gaseosa que llena la siguiente botella, ¿qué diría?
- Con base en los resultados del inciso a) al c), explique por qué es importante realizar un diagrama de series de tiempo y no sólo un histograma como se hizo en el problema 2.27 en la página 43.

**2.44** Los datos en la siguiente tabla representan el número de hogares que utilizaron de forma activa el banco en línea y/o que realizaron pagos en línea de 1995 a 2003. **ONLINEBANKING**

Año	Número de hogares (millones)
1995	0.6
1996	2.5
1997	4.5
1998	7.0
1999	10.5
2000	15.5
2001	22.0
2002	28.0
2003	33.0

Fuente: R.J. Dalton, "In the Mainstream", Newsday, 8 de febrero, 2004, F6-F7.

- Construya un diagrama de series de tiempo para el número de hogares de Estados Unidos que utilizaron activamente el banco en línea y/o pagaron sus cuentas en línea.
- ¿Qué patrón, si lo hay, está presente en estos datos?
- Si tuviera que hacer una predicción en el número de hogares de E.U. que activamente usan el banco en línea y/o hicieron pagos en línea en 2004, ¿qué predeciría?

**2.45** Los datos de la siguiente tabla representan el promedio de espectadores de televisión (excluyendo las televisoras locales) por juego (en millones) para la National Football League (NFL), la National Basketball Association (NBA), la Major League Baseball (MLB) y la National Hockey League (NHL). **SPORT-STV**

Año	NFL	NBA	MLB	NHL
1995	19.6	10.6	15.9	3.6
1996	18.5	10.2	9.8	3.2
1997	17.4	10.8	10.4	2.4
1998	18.1	7.8	9.4	2.6
1999	18.3	7.2	10.0	3.3
2000	17.0	6.7	7.7	2.8
2001	16.9	6.8	9.8	3.1
2002	18.6	5.8	8.9	2.6

Fuente: S. Fatsis, "Salaries, Promos, and Flying Solo", The Wall Street Journal, 9 de febrero, 2004, R.4.

- Para cada uno de los cuatro deportes realice un diagrama de series de tiempo.
- ¿Qué patrón, si lo hay, está presente en estos datos?
- Si tuviera que hacer una predicción del número de espectadores para cada deporte en 2003, ¿qué diría?

## 2.6 USO INADECUADO DE GRÁFICAS Y CONSIDERACIONES ÉTICAS

Las buenas gráficas revelan lo que los datos transmiten. Por desgracia, muchas gráficas presentadas tanto en periódicos como en revistas, así como otras que se desarrollan con el Asistente gráfico de Excel, son incorrectas, engañosas o innecesariamente complicadas, tanto, que nunca deberían utilizarse. Para ilustrar esta situación, la primera gráfica presentada se publicó en la revista *Time* como parte de un artículo sobre la creciente exportación de vino de Australia a Estados Unidos.

**FIGURA 2.17**

Muestra "impropia" de la exportación de vino australiano a Estados Unidos en millones de galones.

Fuente: Adaptado de S. Watterson, "Liquid Gold—Australians Are Changing the World of Wine. Even the French Seem Grateful", Time, 22 de noviembre, 1999, 68.



En la figura 2.17, el icono de la copa de vino que representa los 6.77 millones de galones para 1997 no parece tener casi el doble de tamaño del icono de la copa de vino que representa los 3.67 millones de galones para 1995; tampoco el icono de la copa de vino que representa los 2.25 millones de galones para 1992 parece tener el doble de tamaño del icono de copa de vino que representa 1.04 millones de galones para 1989. La razón para esto, en parte, es que el icono tridimensional de la copa de vino se utiliza para representar las dos dimensiones de exportación y tiempo. Aunque la presentación de la copa de vino puede atraer la vista, los datos deberían presentarse en una tabla de resumen o en un diagrama de series de tiempo.

Además del tipo de distorsión creada por los iconos de la copa de vino en la gráfica de la revista *Time* que muestra la figura 2.17, el uso impropio de los ejes vertical y horizontal lleva a distorsiones. La figura 2.18 en la página 52 presenta otra gráfica usada en el mismo artículo de *Time*.

Existen varios problemas graves en la gráfica. Primero, no hay punto cero en el eje vertical. Segundo, la superficie en acres de 135,326 para el periodo de 1949 a 1950 está trazada por arriba de la superficie en acres de 150,300 para 1969 a 1970. Tercero, no es obvio que la diferencia entre 1979 a 1980 y de 1997 a 1998 (71,569 acres) es aproximadamente tres y media veces la diferencia entre 1979-1980 y 1969-1970 (21,775 acres). Cuarto, no hay valores escalares en el eje horizontal. Los años están trazados junto a los totales de la superficie en acres, no en el eje horizontal. Quinto, los valores para la dimensión del tiempo no están espaciados de forma apropiada a lo largo del eje horizontal. El valor para 1979-1980 es mucho más cercano al de 1990 que a 1969-1970.

**FIGURA 2.18**

Muestra “impropia” de la cantidad de terreno plantado con uvas para la industria vitivinícola.

Fuente: Adaptado de S. Watterson. “Liquid Gold—Australians Are Changing the World of Wine. Even the French Seem Grateful”, Time, 22 de noviembre, 1999, 68-69.



Otros tipos de muestras que atraen la vista y que vemos generalmente en las revistas y periódicos, a menudo incluyen información innecesaria y sólo generan confusión. La figura 2.19 representa una de estas muestras. La gráfica ilustra los productos con la mayor participación de mercado en la industria de las bebidas gaseosas en 1999. Esto genera mucha confusión, aunque la gráfica está diseñada para mostrar las diferencias en la participación de mercado entre las bebidas gaseosas. Las ilustraciones del burbujeo para cada una de las bebidas ocupan mucho espacio de la gráfica respecto a los datos. La misma información podría transmitirse con una gráfica de barras o con una gráfica de pastel.

**FIGURA 2.19**

Diagrama de la participación de mercado de las bebidas gaseosas en 1999.

Fuente: Adaptado de Anne B. Carey y Sam Ward, “Coke Still Has Most Fizz”, USA Today, 10 de mayo, 2000, 1B.



Algunas directrices para desarrollar buenas gráficas son las siguientes:

- La gráfica no debe distorsionar los datos.
- La gráfica no debería contener adornos innecesarios (algunas veces denominados **basura gráfica**).
- Cualquier gráfica de dos dimensiones debe contener una escala para cada eje.
- La escala del eje vertical debe comenzar en cero.
- Todos los ejes deben estar adecuadamente rotulados.
- La gráfica debe tener un título.
- Debe usarse una gráfica lo más sencilla posible para un conjunto de datos.

Una de las fuentes más grandes de gráficas impropias es el Asistente gráfico de Excel. La figura 2.20 representa la ventana de diálogo del paso 1 del Asistente para gráficos. Es posible elegir entre columna, barra, línea, pastel y área de gráfica, así como tipos de gráficas más complicadas como anillos, radial, superficie, burbujas, cotizaciones, cilíndrico, cónico y piramidal. Estas gráficas más complejas deben usarse sólo ocasionalmente pues son más difíciles de interpretar que las gráficas simples que abarcamos en este capítulo.

**FIGURA 2.20**

Ventana de diálogo del paso 1 del Asistente para gráficos de Excel.



La mayoría de los ejemplos del uso inadecuado de las gráficas son resultado de no poner atención en las directrices para crear buenas gráficas. Sin embargo, surgen problemas éticos cuando las gráficas se construyen para mal informar a propósito al lector. Sea como fuere, uno debe extremar precauciones al tratar de obtener conclusiones de las gráficas que se desvían de las directrices mencionadas en este capítulo.

## PROBLEMAS PARA LA SECCIÓN 2.6

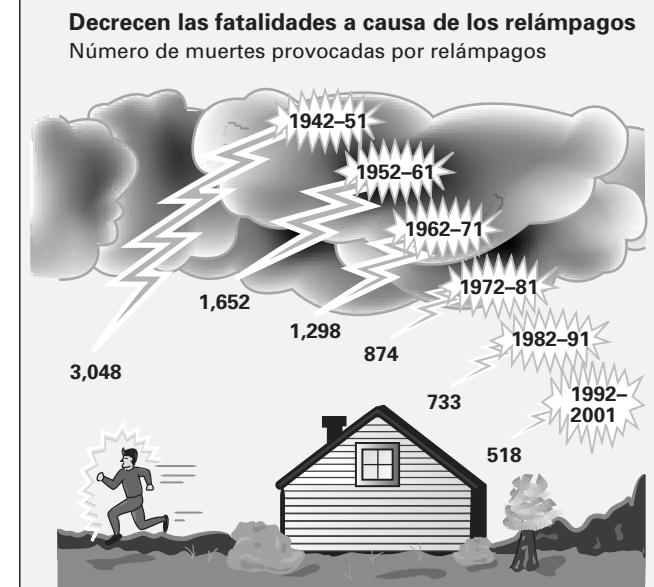
### Aplicación de conceptos

**2.46 (Proyecto estudiantil)** Traiga a clase una gráfica de un periódico o de una revista, que considere que no representa adecuadamente una variable numérica. Preséntela a su maestro con comentarios sobre por qué cree que es inadecuada. ¿Considera que la intención de la gráfica es engañar a propósito al lector? También prepárese para hacer un comentario al respecto en clase.

**2.47 (Proyecto estudiantil)** Traiga a clase una gráfica de un periódico o de una revista, que considere que no representa adecuadamente una variable categórica. Preséntela a su maestro con comentarios sobre por qué cree que sea inadecuada. ¿Considera que la intención de la gráfica es engañar a propósito al lector? También prepárese para hacer un comentario al respecto en clase.

**2.48 (Proyecto estudiantil)** Traiga a clase una gráfica de un periódico o de una revista, que crea que contenga demasiados adornos innecesarios (por ejemplo, basura gráfica) que oscurezcan el mensaje que transmiten los datos. Preséntela a su maestro con comentarios sobre por qué cree que sea inadecuada. También prepárese para hacer un comentario al respecto en clase.

**2.49** La siguiente muestra visual contiene una gráfica sobredimensionada que apareció en el *USA Today* y que se refiere a las muertes a causa de los relámpagos en Estados Unidos.



Fuente: Adaptado de *USA Today*, 12 de noviembre, 2002.

- Describa por lo menos una característica positiva de esta muestra visual.
- Describa por lo menos una característica negativa de esta muestra visual.
- Rediseñe la gráfica usando las directrices mencionadas en la página 52.

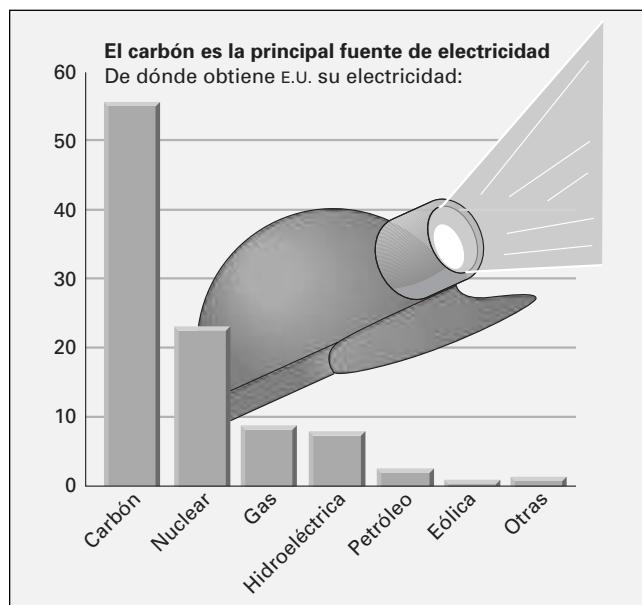
**2.50** La siguiente muestra visual se refiere al tamaño relativo de los departamentos de policía en las ciudades más grandes de Estados Unidos, que apareció en el *USA Today*:



Fuente: Adaptado de *USA Today*, febrero de 2000.

- Indique una característica de esta gráfica que viole los principios de las gráficas bien hechas.
- Diseñe una gráfica alternativa para los datos proporcionados en esta figura.

**2.51** La siguiente muestra visual indica la fuente de la electricidad en Estados Unidos, y apareció en *USA Today*:



Fuente: Adaptado de *USA Today*, 30 de enero, 2002.

## RESUMEN

Como se observa en la tabla 2.18, este capítulo trató acerca de la presentación de datos. Usted ha usado diferentes tablas y gráficas para obtener conclusiones acerca de las compras en línea, el costo de las comidas en restaurantes en la ciudad y en los suburbios, y del conjunto de fondos de inversión que se pre-

- Describa por lo menos una característica positiva de esta muestra visual.
- Describa por lo menos una característica negativa de esta muestra visual.
- Rediseñe la gráfica usando las directrices mencionadas en la página 52.

**2.52** Un artículo publicado en *The New York Times* (Donna Rosato, "Worried about the Numbers? How about the Charts?" *The New York Times*, 15 de septiembre, 2002, Business 7) dio cuenta de una investigación sobre los reportes anuales de las corporaciones, que realizó la profesora Deanna Oxender Burgess de la Florida Gulf Coast University. La profesora Burgess encontró que aun la más mínima distorsión en una gráfica cambia la percepción que el lector tiene de la información. El artículo mostró información sobre el reporte anual de Zale Corporation y expuso la forma en que se exageraron los resultados.

Acuda a Internet o a la biblioteca y estudie el más reciente reporte anual de una corporación. Localice por lo menos una gráfica en el reporte que considere que necesita mejorarse y elabore la gráfica mejorada. Explique por qué cree que su gráfica es mejor que la del reporte anual.

**2.53** Las figuras 2.1, 2.3 y 2.6, en las páginas 23, 24 y 27, consisten en una gráfica de barras, una gráfica de pastel y un diagrama de Pareto para describir los datos de las compras en línea.

- Use el Asistente para gráficos de Excel para desarrollar una gráfica de anillo, una gráfica cónica y una gráfica piramidal para los datos de las compras en línea.
- ¿Qué gráficas prefiere: la gráfica de barras, la gráfica de pastel y el diagrama de Pareto, o la gráfica de anillo, la gráfica cónica y la gráfica piramidal? Explique su respuesta.

**2.54** Las figuras 2.2 y 2.4 de las páginas 24 y 25 consisten en una gráfica de barras y una gráfica de pastel para el nivel de riesgo de los datos de fondos de inversión. **MUTUALFUNDS2004**

- Utilice el Asistente para gráficos de Excel para desarrollar una gráfica de anillo, una gráfica cónica y una gráfica piramidal para el nivel de riesgo de los fondos de inversión.
- ¿Qué gráficas prefiere: la gráfica de barras y la gráfica de pastel o la gráfica de anillo, la gráfica cónica y la gráfica piramidal? Explique su respuesta.

sentaron en el escenario de "Uso de la estadística" al inicio del capítulo. Ahora que ha estudiado tablas y gráficas, en el capítulo 3 aprenderá acerca de una variedad de medidas descriptivas numéricas útiles para el análisis y la interpretación de los datos.

**TABLA 2.18**

Mapa para seleccionar tablas y gráficas.

Tipo de análisis	Tipo de datos	
	Numéricos	Categóricos
Tabular, organizar y presentar gráficamente los valores de la variable	Arreglo ordenado, diagrama de tallo y hojas, distribución de frecuencias, distribución de frecuencias relativas, distribución de porcentaje, distribución acumulativa, histograma, polígono, polígono de porcentaje acumulado ( <b>secciones 2.2 y 2.3</b> )	Tabla de resumen, gráfica de barras, gráfica de pastel, diagrama de Pareto ( <b>sección 2.1</b> )
Presentar gráficamente la relación entre dos variables	Diagrama de dispersión, Series de tiempo ( <b>sección 2.5</b> )	Tabla de contingencia, gráfica de barras agrupadas ( <b>sección 2.4</b> )

## CONCEPTOS CLAVE

Agrupaciones o clases 33	Distribución de porcentaje acumulado	Diagrama de series de tiempo 48
Amplitud de un intervalo de clase 33	35	Polígono de porcentaje 39
Arreglo ordenado 30	Distribución de porcentajes 34	Polígono de porcentaje acumulado 40
Basura gráfica 52	Gráfica de barras 23	Principio de Pareto 25
Celdas 43	Gráfica de barras agrupadas 45	Punto medio de la clase 33
Diagrama de tallo y hoja 30	Gráfica de pastel 24	Rango 33
Diagrama de dispersión 47	Histograma 37	Tabla de datos bivariados 43
Diagrama de Pareto 25	Límites de clase 33	Tabla de contingencia 43
Distribución de frecuencia relativa 34	Ojiva (polígono de porcentaje acumulado) 40	Tabla de resumen 22
Distribución de frecuencias 32		

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**2.55** ¿Cómo es que los histogramas y los polígonos difieren con respecto a su construcción y uso?

**2.56** ¿Cuándo realizaría una tabla de resumen?

**2.57** ¿Cuáles son las ventajas y/o desventajas del uso de la gráfica de barras, la gráfica de pastel o el diagrama de Pareto?

**2.58** Compare y contraste la gráfica de barras para datos categóricos contra el histograma para datos numéricos.

**2.59** ¿Cuál es la diferencia entre el diagrama de series de tiempo y el diagrama de dispersión?

**2.60** ¿Por qué se dice que la característica principal del diagrama de Pareto es que permite separar a lo “poco vital” de lo “mucho trivial”?

**2.61** ¿Qué porcentaje de fracasos pueden ayudar a interpretar los resultados encontrados en una tabla de contingencia?

### Aplicación de conceptos

Puede resolver manualmente los problemas 2.62 a 2.74 o usando Excel, Minitab o SPSS. Le recomendamos usar Excel, Minitab o SPSS para resolver los problemas 2.75 a 2.85.

**2.62** Los datos en la parte superior de la página 56 representan el desglose del precio de un nuevo libro de texto.

Categorías de ingresos	Porcentaje
Editor	64.8
Costos de producción	32.3
Marketing y promoción	15.4
Costos administrativos e impuestos	10.0
Ganancia después de impuestos	7.1
Librería	22.4
Salarios y prestaciones de empleados	11.3
Operaciones	6.6
Ganancias antes de impuestos	4.5
Autor	11.6
Flete	1.2

Fuente: T. Lewin, "When Books Break the Bank", The New York Times, 16 de septiembre, 2003, B1, B4.

- Realice una gráfica de barras, una gráfica de pastel y un diagrama de Pareto para las cuatro categorías de editor, librería, autor y flete.
- Elabore un diagrama de Pareto usando las cuatro subcategorías de editor y las tres subcategorías de librería junto con las categorías de autor y flete.
- Con base en los resultados de los incisos a) y b), ¿a qué conclusiones se llega con relación a quién obtiene las ganancias por la venta de los nuevos libros de texto? ¿Le sorprenden estos resultados? Explique su respuesta.

**2.63** Los siguientes datos representan la participación de mercado para la reparación de automóviles y de camiones ligeros en 1992 y 2002.

Fuente	1992 Porcentaje	2002 Porcentaje
Especialistas foráneos	3.9	6.0
Tiendas de refacciones		
con espacios de servicio	7.3	6.4
Especialistas en reparación	12.7	16.2
Estaciones de servicio, talleres	39.1	29.5
Tiendas de neumáticos	8.1	8.9
Distribuidor de vehículos	21.6	26.6
Otros	7.3	6.4

Fuente: A. Frangos, "Corner Garages Battle Dealers to Fix Your Car", The Wall Street Journal, 3 de junio, 2003, B1, B4.

- Elabore una gráfica de barras, una gráfica de pastel y un diagrama de Pareto para cada año.
- Realice una gráfica de barras agrupada para la participación de mercado en 1992 y 2002.
- Con base en los resultados de los incisos a) y b), ¿qué cambios en la participación de mercado ocurrieron entre 1992 y 2002?

**2.64** Los siguientes datos representan cómo realizaron sus pagos los consumidores en las tiendas en 1999, 2001 y 2003.

Tipo de de pago	1999 Porcentaje	2001 Porcentaje	2003 Porcentaje
Efectivo	39	33	32
Cheque	18	18	15
Débito	21	26	31
Crédito	22	21	21
Otros	0	2	1

Fuente: M. Ingebretsen y M. Ballinger, "Charge It", The Wall Street Journal, 9 de febrero, 2004, R2.

- Construya una gráfica de barras agrupadas para las formas de pago en 1999, 2001 y 2003.
- Con base en los resultados del inciso a), ¿qué cambios en las formas de pago ocurrieron en 1999, 2001 y 2003?

**2.65** Los siguientes datos representan el consumo per cápita de bebidas (en galones) vendidas en tiendas minoristas durante 1998, 2000 y 2002.

Tipo de bebida	1998 Consumo	2000 Consumo	2002 Consumo
Agua embotellada	2.5	4.1	6.7
Lácteo/otros	0.3	0.3	0.3
Jugos	3.1	3.7	4.0
Refrescos	54.0	53.0	52.5
Bebidas energéticas	1.9	2.2	2.5
Té	1.9	2.0	1.9
Total	63.7	65.3	67.9

Fuente: T. Howard, "Coke, Pepsi Sales Up, but Core Colas Flat", USA Today, 21 de julio, 2003, 3B.

- Elabore una tabla de resumen de porcentajes para los tipos de bebida por año.
- Para cada año, elabore una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.
- Elabore una gráfica de barras agrupadas de la participación de mercado de los tipos de bebidas en 1998, 2000 y 2002.
- Con base en los resultados de los incisos a) a c), ¿qué cambios ocurrieron en el mercado entre 1998 y 2002?

**2.66** Brasil es el segundo país consumidor de café en el mundo. A diferencia de la mayoría de los mercados donde unas cuantas corporaciones dominan el tostado y la venta del café, en ese país están activos más de 2,000 pequeños tostadores de café. La corporación Sara Lee se ha convertido en el vendedor de café al detalle líder en Brasil, al adquirir varios tostadores de café (Miriam Jordan, "Sara Lee Wants to Percolate through all Brasil", The Wall Street Journal, 8 de mayo, 2002, A14). En la página 57 se presentan los datos de las siete naciones más consumidoras de café y el desglose de los líderes del mercado en Brasil.

**Principales países consumidores de café en 2000**

País	Consumo (en millones de bolsas de 60 kg)
Estados Unidos	18.6
Brasil	12.8
Alemania	9.2
Japón	6.7
Francia	5.4
Holanda	1.8
Finlandia	0.9

Fuente: The Wall Street Journal.

**Marcas líderes de café en Brasil**

Marca	Participación de mercado
Marcas pertenecientes a Sara Lee	27.6%
Nescafé	6.1%
Tres Coracoes	4.8%
Melitta	4.0%
Todas las demás	57.5%

Fuente: The Wall Street Journal.

- Construya una gráfica para los datos relacionados con los países que consumen más café. ¿Qué tipo de gráfica es la más apropiada? Explique por qué.
- Construya una gráfica para los datos referentes a la participación de mercado en Brasil. ¿Qué tipo de gráfica es la más adecuada? Explique por qué.

**2.67** Los siguientes datos representan las reservas convencionales probadas de petróleo en miles de millones de barriles, subdivididos por región y país.

Región y país	Reservas convencionales probadas (en miles de millones de barriles)	Reservas
Norteamérica	54.8	
México		28.3
E.U.		21.8
Canadá		4.7
Centro y Sudamérica	95.2	
Venezuela		76.9
Brasil		8.1
Otros en Centro y Sudamérica		10.2
Europa Occidental	17.2	
Noruega		9.5
Inglaterra		5.0
Otros en Europa Occidental		2.7
África	74.9	
Libia		29.5
Nigeria		22.5
Argelia		9.2
Angola		5.4
Otros en África		8.3

Región y país	Reservas convencionales probadas (en miles de millones de barriles)	Reservas
Medio Oriente	683.6	
Arabia Saudita		259.2
Irak		112.5
Emiratos Árabes Unidos		97.8
Kuwait		94.0
Irán		89.7
Qatar		13.2
Omán		5.5
Otros en Medio Oriente		11.7
Lejano Oriente y Oceanía	44.0	
China		24.0
Indonesia		5.0
India		4.7
Otros en Lejano Oriente y Oceanía		10.3
Europa Oriental y ex URSS	59.0	
Rusia		48.6
Kazajstán		5.4
Otros en Europa Oriental y ex URSS		5.0

Fuente: Departamento de Energía de Estados Unidos.

Usando el conjunto de países:

- Construya una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.

Usando el conjunto de regiones:

- Realice una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.
- ¿Qué método gráfico cree que refleja mejor estos datos?
- Con base en los resultados de los incisos a) y b), ¿a qué conclusiones se llega respecto a las reservas convencionales de petróleo probadas para los diferentes países y regiones?

**2.68** Al analizar las repercusiones de los ataques del 11 de septiembre de 2001, los especialistas del National Center for Health Statistics, se enfocaron en su habilidad de localizar y clasificar a las víctimas del terrorismo (E. Weinstein, "Tracking Terror's Rising Toll", *The Wall Street Journal*, 25 de enero, 2002, A13). Los siguientes datos representan las muertes provocadas por el terrorismo en Estados Unidos entre 1990 y 2001 y también las muertes en Estados Unidos provocadas por diversas causas.

Año	Muertes provocadas por el terrorismo en Estados Unidos
1990	0
1991	0
1992	0
1993	6
1994	1
1995	169
1996	2
1997	0
1998	1
1999	3
2000	0
2001	2,717

Causa	Muertes en miles
Humo y fuego	3.3
Ahogamiento accidental	3.3
Muertes inducidas por el alcohol	18.5
Enfermedad de Alzheimer	49.0
Asalto con armas de fuego	10.4
Asalto sin armas de fuego	5.7
Asma	4.4
Cáncer	551.8
Embolia y enfermedades relacionadas	166.0
Enfisema	16.9
Diabetes	68.7
Enfermedades cardíacas	710.0
Caídas	12.0
VIH	14.4
Influenza y neumonía	67.0
Lesiones en el trabajo	5.3
Accidentes en vehículos	41.8
Suicidio	28.3
Muertes relacionadas con las drogas	15.9

Fuente: Federal Bureau of Criminal Justice Statistics, National Center for Health Statistics, National Highway Transportation, Safety Administration, Departamento de Defensa de Estados Unidos.

- a. Diseñe un diagrama de series de tiempo para las muertes provocadas por el terrorismo en territorio estadounidense. ¿Hay algún patrón en las muertes provocadas por el terrorismo en territorio estadounidense entre 1990 y 2001?

Para las muertes en miles de personas por diferentes causas:

- b. Elabore una gráfica de barra, una gráfica de pastel y un diagrama de Pareto.  
 c. ¿Qué método gráfico es el mejor para describir estos datos?  
 d. Con base en los resultados del inciso c), ¿a qué conclusiones se llega respecto a las muertes en Estados Unidos provocadas por diversas causas?

**2.69** El dueño de un restaurante que sirve platillos estilo Continental estaba interesado en estudiar los patrones de demanda para el periodo de viernes a domingo (fin de semana). Guardó registros que indicaban el número de órdenes para cada tipo de platillo. Los datos son los siguientes:

Tipo de platillo	Número servido
Res	187
Pollo	103
Pato	25
Pescado	122
Pasta	63
Mariscos	74
Ternera	26

- a. Elabore una tabla de resumen para los tipos de platillos ordenados.

- b. Elabore una gráfica de barras, una gráfica de pastel y un diagrama de Pareto para los tipos de platillos ordenados.  
 c. ¿Prefiere un diagrama de Pareto o una gráfica de pastel para estos datos? ¿Por qué?  
 d. ¿A qué conclusiones llegará el dueño del restaurante en relación con la demanda de los diferentes tipos de platillos?

**2.70** Suponga que el dueño del restaurante del problema 2.69 también estaba interesado en estudiar la demanda de postres para el mismo periodo. Decidió que otras dos variables, junto con la variable de si se ordena o no el postre, debían estudiarse: el género del individuo y si ordenaba el platillo de res. Los resultados fueron los siguientes:

GÉNERO			
ORDENÓ POSTRE	Masculino	Femenino	Total
Sí	96	40	136
No	224	240	464
Total	320	280	600
PLATILLO DE RES			
ORDENÓ POSTRE	Sí	No	Total
Sí	71	65	136
No	116	348	464
Total	187	413	600

Para cada una de las dos tablas de contingencia:

- a. Realice una tabla de porcentajes de fila, de porcentajes de columna y de porcentajes totales.  
 b. ¿Qué tipo de porcentaje (fila, columna o total) cree que es el más informativo para cada género? ¿Para el platillo de res? Explique por qué.  
 c. ¿Qué conclusiones obtendrá el dueño del restaurante respecto al patrón de postres ordenados?

**2.71** Un artículo publicado en *The New York Times* (William McNulty y Hugh K. Truslow, "How it Looked Inside the Booth", *The New York Times*, 6 de noviembre, 2002) aportó los siguientes datos para el método de registro de votos en 1980, 2000 y 2002, divididos en los porcentajes de ciudades en Estados Unidos que emplearon cada método y los porcentaje de votantes registrados que utilizaron cada uno. Los resultados son los siguientes:

MÉTODO	PORCENTAJE DE CIUDADES QUE LO USARON		
	1980	2000	2002
Tarjetas perforadas	18.5	18.5	15.5
Máquinas con palanca	36.7	14.4	10.6
Papeletas	40.7	11.9	10.5
Escaneo óptico	0.8	41.5	43.0
Electrónico	0.2	9.3	16.3
Mixto	3.1	4.4	4.1

MÉTODO	PORCENTAJE DE VOTANTES REGISTRADOS QUE LO USARON		
	1980	2000	2002
Tarjetas perforadas	31.7	31.4	22.6
Máquinas con palanca	42.9	17.4	15.5
Papeletas	10.5	1.5	1.3
Escaneo óptico	2.1	30.8	31.8
Electrónico	0.7	12.2	19.6
Mixto	12.0	6.7	9.3

- Realice gráficas de pastel para cada año para los porcentajes de las ciudades y los porcentajes de los votantes registrados que usaron los diversos métodos.
- Realice gráficas de barras agrupadas por año para los porcentajes de ciudades y los porcentajes de votantes registrados que usaron los diferentes métodos.
- ¿Qué tipo de muestra gráfica es más útil para representar los datos? Explique por qué.
- ¿Qué diferencias existen en los resultados para las ciudades y para los votantes registrados?

**2.72** En el verano del año 2000, se registró un mayor número de reclamaciones sobre la garantía de los neumáticos Firestone vendidos en la Ford SUV, lo que provocó que tanto la Firestone como la Ford emitieran un importante comunicado. Un análisis de los datos de quejas sobre la garantía ayudó a identificar qué modelos debían atender. El desglose de 2,504 quejas de garantía basadas en el tamaño del neumático se presenta en la siguiente tabla:

Rodada del neumático	Quejas de garantía
23575R15	2,030
311050R15	137
30950R15	82
23570R16	81
331250R15	58
25570R16	54
Otros	62

Fuente: Robert L. Simison, "Ford Steps Up Recall without Firestone", The Wall Street Journal, 14 de agosto, 2000, A3.

Las 2,030 quejas de garantía para los neumáticos 23575R15 pueden clasificarse en los modelos ATX y Wilderness. El tipo de incidente que llevó a la queja de garantía, por tipo de modelo, se resume en la siguiente tabla.

Incidente	Quejas de garantía para el modelo ATX	Quejas de garantía para el modelo Wilderness
Separación de la banda de rodadura	1,365	59
Ponchadura	77	41
Otro/desconocido	422	66
Total	1,864	166

Fuente: Robert L. Simison, "Ford Steps Up Recall without Firestone", The Wall Street Journal, 14 de agosto, 2000, A3.

**a.** Elabore un diagrama de Pareto para el número de quejas de garantía por la rodada del neumático. ¿Qué rodada del neumático recibió el mayor número de quejas?

**b.** Elabore una gráfica de pastel que muestre el porcentaje del número total de quejas de garantía para los neumáticos 23575R15 que vienen en el modelo ATX y en el modelo Wilderness. Interprete la gráfica.

**c.** Realice un diagrama de Pareto para el tipo de incidente causado en la reclamación de garantía para el modelo ATX. ¿Existe un cierto tipo de incidente que cause la mayoría de las quejas?

**d.** Diseñe un diagrama de Pareto para el tipo de incidente que causa la queja de garantía para el modelo Wilderness. ¿Existe un cierto tipo de incidente que cause la mayoría de las quejas?

**2.73** Una de las principales medidas de calidad del servicio que brinda cualquier organización es la velocidad con la que responde a las quejas del consumidor. Una gran tienda, departamental que vende muebles, pisos y alfombras, ha tenido una gran expansión en los últimos años. El departamento de pisos, en especial, ha crecido de 2 cuadrillas de instalación a un supervisor de instalación, un medidor y 15 cuadrillas de instalación. Durante un año reciente, la empresa recibió 50 quejas referentes a la instalación de alfombras. Los siguientes datos representan el número de días entre la recepción de la queja y su solución. **FURNITURE**

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

**a.** Elabore una distribución de frecuencia y una distribución de porcentaje.

**b.** Trace un histograma y un polígono de porcentaje.

**c.** Construya una distribución de porcentaje acumulado y trace una ojiva (polígono de porcentaje acumulado).

**d.** Con base en los resultados de los incisos a) a c), si tuviera que informar al presidente de la empresa cuánto tiene que esperar el cliente para que su queja se resuelva, ¿qué le diría? Explique su respuesta.

**2.74** Los datos del archivo **PIZZA** representan el costo por rebanada en dólares, el número de calorías por rebanada y la cantidad de grasa en gramos por rebanada para una muestra de 36 productos de pizza.

Fuente: "Frozen Pizza on the Rise", Copyright © 2002 por Consumer Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, enero de 2002, 40-41.

**a.** Elabore distribuciones de frecuencias y distribuciones de porcentaje para grasa, costo y calorías.

**b.** Construya histogramas y polígonos de porcentaje para grasa, costo y calorías.

- c. Elabore distribuciones de porcentaje acumulado y trace ojivas (polígonos de porcentaje acumulado) para grasa, costo y calorías.
- d. Elabore diagramas de dispersión para el costo y las calorías, costo y grasa, y calorías y grasa.
- e. Con base en el inciso a) al d), ¿a qué conclusiones llega acerca del costo, grasas y calorías de estos productos de pizza?

**2.75** Un artículo en *Quality Engineering* examinó la viscosidad (resistencia al flujo) de un producto químico producido en partidas. Tome en cuenta que la viscosidad del químico necesita estar entre 13 y 18 para cumplir con las especificaciones de la empresa. Los datos para las 120 partidas están en los datos del archivo **CHEMICAL**.

Fuente: D.S. Holmes y A.E. Mergen, "Parabolic Control Limits for the Exponentially Weighted Moving Average Control Charts", *Quality Engineering*, vol. 4 (1992), 487-495.

- a. Elabore un arreglo ordenado.
- b. Construya una distribución de frecuencia y una distribución de porcentaje.
- c. Elabore un histograma de porcentaje.
- d. ¿Qué porcentaje de las partidas está dentro de las especificaciones de la empresa?

**2.76** Estudios que realizó un fabricante de tablillas de asfalto "Boston" y "Vermont" demuestran que el peso del producto es el factor principal en la percepción que el consumidor tiene de la calidad. Más aún, el peso representa la cantidad de material usado, por lo que es muy importante para la empresa desde el punto de vista del costo. La última etapa de la línea de ensamblaje empaca las tablillas antes de que los paquetes se coloquen en paletas de madera. Una vez que la paleta está llena (las paletas de casi todas las marcas tienen capacidad para sostener 16 cuadros de tablillas), se pesa y se registra este dato. La empresa espera que las paletas de tablillas de su marca "Boston" pesen por lo menos 3,050 libras pero menos de 3,260. Para las tablillas de la marca "Vermont", las paletas deben pesar por lo menos 3,600 libras pero menos de 3,800. El archivo de datos **PALLET** contiene los pesos (en libras) de una muestra de 368 paletas de tablillas "Boston" y 330 paletas de tablillas "Vermont".

- a. Para las tablillas "Boston", construya una distribución de frecuencias y una distribución de porcentaje con ocho intervalos de clase usando como límites de clase: 3,015, 3,050, 3,085, 3,120, 3,155, 3,190, 3,225, 3,260 y 3,295.
- b. Para las tablillas "Vermont", construya una distribución de frecuencias y una distribución de porcentaje con siete intervalos de clase, usando como límites de clase: 3,550, 3,600, 3,650, 3,700, 3,750, 3,800, 3,850 y 3,900.
- c. Elabore histogramas para las tablillas "Boston" y para las tablillas "Vermont".
- d. Comente acerca de la distribución del peso de las paletas para las tablillas "Boston" y "Vermont". Asegúrese de identificar el porcentaje de paletas que se encuentran por debajo y por encima del peso.

**2.77** ¿Regalar artículos promocionales incrementa la asistencia a los juegos de la liga de béisbol? Un artículo en *Sport Mar-*

*keting Quarterly* reportó sobre el efecto de las promociones mercadológicas [T.C. Boyd y T.C. Krehbiel, "Promotion Timing in Major League Baseball and the Stacking Effects of Factors that Increase Game Attractiveness", *Sport Marketing Quarterly*, vol. 12 (2003), 173-184]. El archivo de datos **ROYALS** incluye las siguientes variables para el equipo Kansas City Royals durante la temporada de béisbol del año 2002:

GAME = Juegos en casa en el orden en el que se jugaron.

ATTENDANCE = Asistencia pagada al juego.

PROMOTION 1 = Si se realizó una promoción; 0 = si no se realizó ninguna promoción.

- a. Elabore un histograma de porcentaje para la variable de asistencia. Interprete el histograma.
- b. Elabore un polígono de porcentaje para la variable de asistencia. Interprete el polígono.
- c. ¿Qué gráfica prefiere: la del inciso a) o la del inciso b)? Explique.
- d. Elabore una gráfica que contenga dos polígonos de porcentaje para la asistencia: uno para los 43 juegos con promoción y uno para los 37 juegos sin promoción. Compare las distribuciones de las dos asistencias.

**2.78** Los datos para el archivo **PROTEIN** indican las calorías, proteínas, calorías de grasa, calorías de grasa saturada y colesterol para comidas proteínicas populares (carnes rojas frescas, pollo y pescado).

Fuente: Departamento de Agricultura de Estados Unidos.

Para los datos referentes al número de calorías, proteínas, calorías de la grasa, calorías de grasa saturada y cantidad de colesterol para las comidas proteínicas populares:

- a. Elabore una distribución de frecuencias y una distribución de porcentajes.
- b. Elabore un histograma y un polígono de porcentaje.
- c. Construya una distribución de porcentaje acumulado, y elabore un polígono de porcentaje acumulado.
- d. ¿A qué conclusiones llega a partir de estos análisis?

**2.79** Suponga que desea estudiar las características de los modelos de automóviles del año 2002 en términos de las siguientes variables: caballos de potencia, millas por galón, longitud, anchura, requerimiento de ángulo de giro, peso y volumen de carga. **AUTO2002**

Fuente: "The 2002 Cars", Copyright © 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, abril de 2002, 22-71.

Para cada una de estas variables:

- a. Elabore una distribución de frecuencias y una distribución de porcentajes.
- b. Construya un histograma y un polígono de porcentajes.
- c. Elabore una distribución de porcentaje acumulado y trace un polígono de frecuencia acumulada.
- d. ¿A qué conclusiones llega respecto a los automóviles 2002?

**2.80** En relación con las características de los modelos de automóviles del año 2002 **AUTO2002** del problema 2.79,

- a. Construya una tabla de contingencia del tipo de tracción con el tipo de gasolina.
- b. Elabore una gráfica de barras agrupadas del tipo de tracción con el tipo de gasolina.

- c. Con base en los resultados de los incisos a) y b), ¿parece haber relación entre el tipo de transmisión y el tipo de gasolina?

**2.81** Los datos del archivo **STATES** representan los resultados de la encuesta American Community, una muestra de hogares tomados de todos los estados durante el Censo de 2000 de Estados Unidos. Para cada una de las variables de promedio del tiempo en minutos del trayecto al trabajo, porcentaje de casa con ocho o más habitaciones, ingreso promedio del hogar y porcentaje de pagos hipotecarios que realizan los propietarios cuyos costos de casa exceden el 30% de su ingreso:

- Elabore una distribución de frecuencias y una distribución de porcentajes.
- Construya un histograma y un polígono de porcentaje.
- Construya una distribución de porcentaje acumulado y trace un polígono de porcentaje acumulado.
- ¿Qué conclusiones se obtienen de estas cuatro variables con base en los resultados de los incisos a) al c)?

**2.82** La economía del béisbol ha causado una gran controversia desde que los propietarios de los equipos se quejaron de sus pérdidas económicas; los jugadores argumentan que los propietarios están ganando dinero y los fanáticos se quejan de lo caro que resulta asistir a un juego, por lo que prefieren quedarse en casa y ver los partidos por televisión. Además de los datos relacionados con las estadísticas de los equipos para la temporada 2001, el archivo **BB2001** contiene las estadísticas del precio de los boletos de cada equipo; el índice del costo por fanático; ingresos por entrada en una temporada regular; ingresos por televisión, radio y cable local; todos los demás ingresos operativos; la compensación y los beneficios de los jugadores; otros gastos locales y nacionales, así como el ingreso de las operaciones derivadas del béisbol. Para cada una de estas variables,

- Construya una distribución de frecuencias y una distribución de porcentaje.
- Elabore un histograma y un polígono de porcentaje.
- Elabore una distribución de porcentaje acumulado y construya un polígono de porcentaje acumulado.
- Elabore un diagrama de dispersión para predecir el número de juegos ganados en el eje Y, y la compensación y beneficios del jugador en el eje X. ¿Qué conclusiones obtiene a partir de este diagrama de dispersión?
- ¿Qué conclusiones acerca de estas variables se obtienen con base en los resultados de los incisos a) al c)?

**2.83** Los datos en el archivo **AIRCLEANERS** representan el precio, el costo anual de energía y el costo anual de los filtros para los limpiadores de aire de las habitaciones.

- Construya un diagrama de dispersión con el precio en el eje Y, y el costo de energía en el eje X.
- Elabore un diagrama de dispersión con el precio en el eje Y, y el costo de los filtros en el eje X.
- ¿A qué conclusiones se llega sobre el costo de la energía y el costo de los filtros en relación con el de los limpiadores de aire?

Fuente: "Portable Room Air Cleaners", Copyright © 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, febrero de 2002, 47.

**2.84** Los datos del archivo **PRINTERS** representan el precio, la velocidad del texto, costo del texto, tiempo de las fotografías de color y costo de las impresiones de las fotografías a color.

- Elabore un diagrama de dispersión con el precio y la velocidad del texto, costo y costo del texto, precio y tiempo de las fotografías a color, y precio y costo de las fotografías a color.
- Con base en los resultados del inciso a), ¿cree usted que alguna de las otras variables serían útiles para predecir el precio de la impresora? Explique su respuesta.

Fuente: "Printers", Copyright 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con permiso de Consumer Reports, marzo de 2002, 51.

**2.85** El Índice S&P 500 rastrea el movimiento general del mercado bursátil de acuerdo con el precio de las acciones de 500 grandes corporaciones. El archivo **STOCKS2003** contiene los datos semanales para este índice, así como el precio semanal del cierre de las acciones para tres compañías durante 2003. Las variables incluidas son:

WEEK —Semana que termina en una fecha dada.

S & P —Cierre del valor semanal para el Índice S&P 500.

SEARS —Cierre semanal del precio de las acciones de Sears.

TARGET —Cierre semanal del precio de las acciones de Target.

SARA LEE —Cierre semanal del precio de las acciones de Sara Lee.

Fuente: [finance.yahoo.com](http://finance.yahoo.com)

- Elabore un diagrama de series de tiempo para los valores de cierre semanal del Índice S&P 500 Sears, Roebuck and Company, Target Corporation y Sara Lee.
- Explique cualquier patrón que detecte en la gráfica.
- Escriba un resumen de sus hallazgos.

**2.86 (Proyecto de clase)** Permita que cada alumno de su clase responda a la pregunta "¿Qué bebida refrescante prefiere?", para que el maestro registre los resultados en una tabla de resumen.

- Convierta los datos a porcentajes para construir un diagrama de Pareto.
- Analice sus hallazgos.

**2.87 (Proyecto de clase)** Permita que cada alumno de su clase sea clasificado con base en su género (masculino, femenino) y estado de empleo actual (sí, no) para que el maestro registre los resultados.

- Construya una tabla con porcentajes de fila o de columna, dependiendo de cuál considere que es más informativo.
- ¿Qué concluiría de este estudio?
- ¿Qué otras variables desearía conocer respecto al empleo con la finalidad de enriquecer sus hallazgos?

### Reportes escritos de los ejercicios

**2.88** De acuerdo con los resultados del problema 2.76 en la página 60 respecto al peso de las tablillas "Boston" y "Vermont", escriba un reporte que evalúe si el peso de las paletas de los dos tipos de tablillas es congruente con las expectativas de la empresa. Asegúrese de incluir tablas y gráficas en el reporte.

**2.89** De acuerdo con los resultados del problema 2.72 en la página 59 respecto a las quejas de la garantía de los neumáticos Firestone, escriba un reporte que evalúe las quejas de garantía de los neumáticos Firestone vendidos en los modelos Ford SUV. Asegúrese de incluir tablas y gráficas en el reporte.



## PROYECTO EN EQUIPO

Los datos del archivo **MUTUALFUNDS2004** contienen información respecto a 12 variables de una muestra de 121 fondos de inversión. Las variables son:

Fund —El nombre del fondo de inversión.

Category —Tipo de acciones que comprende el fondo de inversión: de gran capital, capital medio, capital reducido.

Objective —Objetivo de las acciones que comprenden el fondo de inversión: crecimiento o valor.

Assets —Activos en millones de dólares.

Fees —Cargos por ventas (sí o no).

Expense ratio —Relación entre gastos y activos netos en porcentaje.

2003 Return —Rendimiento de 12 meses en 2003.

Three-year return —Rendimiento anualizado de 2001 a 2003.

Five-year return —Rendimiento anualizado de 1999 a 2003.

Risk —Factor de riesgo de pérdida de los fondos de inversión clasificado como bajo, promedio o alto.

Best quarter —Mejor resultado trimestral 1999 a 2003.

Worst quarter —Peor resultado trimestral 1999 a 2003.

**2.90** Para la variable de relación de gasto:

- Elabore un histograma.
- Construya polígonos de frecuencia para la proporción de gasto para los fondos de inversión que tengan tarifas y para los que no tengan tarifas en la misma gráfica.
- ¿Qué conclusiones acerca de la proporción de gasto se obtienen con base en los resultados de los incisos a) y b)?

**2.91** Para la variable que contiene el rendimiento anualizado de cinco años de 1999 a 2003:

- Elabore un histograma.
- Construya polígonos de frecuencia del rendimiento anualizado de cinco años de 1999 a 2003 para los fondos de inversión de crecimiento y para los fondos de inversión de valor en la misma gráfica.
- ¿A qué conclusiones llega acerca del rendimiento anualizado de cinco años entre 1999 y 2003, con base en los resultados de los incisos a) y b)?

**2.91** Para la variable que contiene el rendimiento anualizado de tres años de 2001 a 2003:

- Elabore un histograma.
- Construya polígonos de frecuencia del rendimiento anualizado de tres años entre 2001 y 2003 para los fondos de inversión de crecimiento y para los fondos de inversión de valor en la misma gráfica.
- ¿A qué conclusiones llega acerca del rendimiento anualizado de tres años entre 2001 y 2003, con base en los resultados de los incisos a) y b)?

## CASO ACTUAL ADMINISTRACIÓN DEL SPRINGVILLE HERALD

Las tarifas de publicidad son una importante fuente de ingresos para cualquier periódico. En un intento de incrementar estos ingresos y de minimizar errores costosos, la administración del *Herald* formó un equipo de trabajo encargado de mejorar el servicio al consumidor en el departamento de publicidad. Revise la colección de datos de la fuerza de trabajo (abra **Ad\_Errors.htm** en la carpeta Springville HeraldCase del CD-ROM que acompaña a este texto o visite el sitio [www.prenhall.com/Herald Case/Ad\\_Errors.htm](http://www.prenhall.com/Herald Case/Ad_Errors.htm)) e identifique los datos que son importantes para describir los problemas del servicio al consumidor. Para cada conjunto de datos que identifique, elabore la gráfica que considere más adecuada para los datos y explique su elección. También sugiera qué otra información respecto a los diferentes tipos de errores sería útil examinar. Ofrezca posibles cursos de acción que el equipo de trabajo o la administración puedan tomar para mejorar el servicio al consumidor.

se/Ad\_Errors.htm) e identifique los datos que son importantes para describir los problemas del servicio al consumidor. Para cada conjunto de datos que identifique, elabore la gráfica que considere más adecuada para los datos y explique su elección. También sugiera qué otra información respecto a los diferentes tipos de errores sería útil examinar. Ofrezca posibles cursos de acción que el equipo de trabajo o la administración puedan tomar para mejorar el servicio al consumidor.

## CASO WEB

*En el escenario de “Uso de la estadística” se le pidió que recabara información que ayudara a elegir buenas opciones de inversión. Las fuentes para tal información incluyen firmas de correduría y consejeros de inversión. Aplique sus*

*conocimientos sobre el uso correcto de las tablas y gráficas en este Caso Web sobre las afirmaciones de previsión y excelencia de un servicio de inversión de Springville.*

Visite el sitio en Internet de StockTout Investing Service en [www.prenhall.com/Springville/StockToutHome.htm](http://www.prenhall.com/Springville/StockToutHome.htm). Revise sus afirmaciones sobre inversión y los datos que las apoyan, y responda a lo siguiente:

1. ¿Cómo afecta la percepción que usted tiene de sus negocios la presentación de la información general acerca de StockTout en esta página?
2. ¿La afirmación de tener más ganadores que perdedores es una reflexión justa y precisa sobre la calidad de su servicio de inversión? Si no lo cree así, ofrezca una presentación alternativa que considere justa y precisa.

3. Los fondos de inversión de StockTout “Big Eight” son parte de la muestra encontrada en el archivo **MUTUALFUNDS2004**. ¿Hay otros datos relevantes en ese archivo que pudieran incluirse en la tabla Big Eight? ¿Cómo alterarían estos nuevos datos su percepción de las afirmaciones de StockTout?

4. StockTout se enorgullece de que los fondos “Big Eight” han ganado valor en los últimos cinco años. ¿Está de acuerdo en que deben estar orgullosos de sus elecciones? ¿Por qué sí o por qué no?

## REFERENCIAS

1. Huff, D., *How to Lie with Statistics* (Nueva York: Norton, 1954).
2. *Microsoft Excel 2003* (Redmond, WA: Microsoft Corporation, 2002).
3. *Minitab for Windows Version 14* (State College; PA: Minitab Inc., 2004).
4. *SPSS ® Base 12.0 Brief Guide* (Upper Saddle River, NJ: Prentice Hall, 2003).
5. Tufte, E.R., *Envisioning Information* (Cheshire, CT: Graphics Press, 1990).
6. Tufte, E.R., *The Visual Display of Quantitative Information*, 2a. ed. (Cheshire, CT: Graphics Press, 2002).
7. Tufte, E.R., *Visual Explanations* (Cheshire, CT: Graphics Press, 1997).
8. Wainer, H., *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot* (Nueva York: Copernicus/Springer-Verlag, 1997).

# Apéndice 2 Uso del software para tablas y gráficas

## A2.1 EXCEL

Puede usar Excel para crear muchas de las tablas y gráficas explicadas en este capítulo. Si no ha leído aún el apéndice 1.2 “Introducción a Excel”, en la página 13, debe hacerlo ahora.

### Tablas de resumen

Use el Asistente para tablas y gráficos dinámicos para generar una tabla de resumen. Si no está familiarizado con las Tablas dinámicas, primero lea “Uso del Asistente para tablas y gráficos dinámicos” (vea el apéndice F). Para generar una tabla de resumen similar a la tabla 2.2 en la página 23, abra la hoja de

trabajo **MUTUALFUNDS2004.xls** en **Datos**. Seleccione **Datos** → **Informe de tablas y gráficos dinámicos** (**Datos** → **Informe de tablas dinámicas** en Excel 97) y haga estas elecciones en las ventanas de diálogo del Asistente de tablas y gráficos dinámicos:

**Paso 1:** Seleccione la **lista Excel** o la opción **Datos** y la opción **Informe de tablas y gráficos dinámicos** (si aparece) y dé clic en **Siguiente**.

**Paso 2:** Ingrese **J1:J122** como **Rango** y dé clic en **Siguiente**.

**Paso 3:** Seleccione la opción **Hoja de cálculo nueva** y dé clic en el botón de **Diseño**.

En la ventana de diálogo de Diseño, primero mueva una copia de la pestaña **Riesgo** al área de **FILA**. Después mueva la segunda copia de la pestaña **Riesgo** al área de **DATOS**, lo que cambia la pestaña a **Conteo de riesgo**. Dé clic en **Aceptar** para regresar a la ventana de diálogo principal del paso 3 y dé clic en el botón de **Opciones** para continuar.

En la ventana de diálogo de Opciones para la tabla dinámica, ingrese un nombre autodescriptivo para la tabla en el cuadro de edición **Nombre** y **0** en el cuadro de edición **Mostrar para celdas vacías**. Dé clic en **Aceptar** para regresar a la ventana de diálogo principal del paso 3.

Dé clic en **Finalizar** en la ventana de diálogo principal del paso 3 para producir la tabla dinámica.

Renombre la nueva hoja de trabajo con un nombre autodescriptivo. (Puede cerrar cualquier barra de herramientas o ventanas que aparezcan sobre la tabla dinámica para mejorar la visualización.)

Para agregar una columna de porcentaje, ingrese **Porcentaje** en la celda **C4** de la nueva hoja de trabajo e ingrese la fórmula **=B5/B\$8** en la celda **C5**. Copie esta fórmula hacia abajo hasta la celda **C7**. Formatee el rango de celda **C5:C7** para mostrar el porcentaje. Ajuste el número de decimales mostrados y el ancho de la columna C si desea generar una tabla similar a la de la figura A2.1.

Conteo de riesgo		
Riesgo	Total	Porcentaje
promedio	46	38%
alto	17	14%
bajo	58	48%
Total	121	

**FIGURA A2.1** Tabla de resumen completa.

O Vea la sección G.1 (Tablas y gráficas de una variable) si desea que el PHStat2 genere una tabla de resumen para usted.

### Gráfica de barras o de pastel

Use el Asistente para gráficos de Excel para generar una gráfica de barra o de pastel. Si no está familiarizado con este asistente, lea antes “Uso del Asistente para gráficos de Excel” (página 15). Primero cree una tabla de resumen de la tabla dinámica. Con la tabla en pantalla, dé clic en una celda fuera de la tabla, seleccione **Insertar → Gráfica** y haga las siguientes entradas en las ventanas de diálogo del Asistente gráfico:

**Paso 1:** Dé clic en **Barra** (para una gráfica de barra) o en **Pastel** en la ventana **Tipos estándar** **Tipo de gráfico** y deje seleccionado el primer **Subtipo de gráfico**. Dé clic en **Siguiente**.

**Paso 2:** Con el cursor parpadeante en la ventana **Rango de datos**, dé clic en la tabla dinámica para que Excel llene la dirección de la tabla dinámica por usted. Dé clic en **Siguiente**.

**Paso 3:** Seleccione las opciones de formato y rótulos de gráfico para la gráfica. (Vea “Uso del Asistente para gráficos de Excel” en la página 15 para sugerencias.) Dé clic en **Siguiente**.

**Paso 4:** Seleccione **En una hoja nueva** y dé clic en **Finalizar**. Si aparecen botones de campo en la gráfica, dé un clic derecho en cualquier botón del campo y seleccione **Ocultar botones de campo gráfico dinámico** del menú corto.

O Vea la sección G.1 (Gráficos y tablas de una variable) si desea que PHStat2 genere para usted un diagrama de Pareto como una gráfica de Excel. (No hay comandos en Excel que generen directamente un diagrama de Pareto.)

### Arreglo ordenado

Organice su hoja de trabajo para que cada variable aparezca en su propia columna, ingrese un título de la variable columna en la fila 1 y los valores para la variable iniciando en la fila 2. (Éste es el formato de los archivos Excel incluidos en el CD-ROM que acompaña este texto.) Seleccione **Datos → Ordenar**. En la ventana de diálogo de Ordenar, seleccione la variable a ordenar de la lista hacia abajo **Ordenar por**. Seleccione el primer botón de la opción **Ascendente** o **Descendente**, deje seleccionado el botón **Encabeza fila** y dé clic en **Aceptar**.

### Diagrama de tallo y hojas

Vea la sección G2 (Diagrama de tallo y hojas) si desea que PHStat2 genere un diagrama de tallo y hojas como gráfica de Excel. (No hay comandos en Excel que generen directamente estos diagramas.)

### Distribuciones de frecuencias e histogramas

Use el Data Analysis Toolpak para crear distribuciones de frecuencias e histogramas. Abra la hoja de trabajo que contenga los datos que deseé resumir. Seleccione **Herramientas → Análisis de datos**. De la lista que aparece en la ventana de diálogo de Análisis de datos, seleccione **Histograma** y dé clic en **Aceptar**. En la ventana de diálogo de Histograma (vea la figura A2.2) in-



**FIGURA A2.2** Ventana de diálogo del histograma análisis de datos.

grese el rango de celda de los datos en **Rango de entrada**. Después seleccione **Rótulos** si está utilizando datos que estén ordenados como los datos en los archivos Excel del CD-ROM que acompaña a este texto. Por último, seleccione **Resultado de la gráfica** y dé clic en **Aceptar**. (Vea la sección G.3 [Histogramas y polígonos] para una explicación sobre el **Bin Range**.)

La distribución de frecuencias y el histograma aparecerán juntos o en una hoja de trabajo separada.

*O* Vea la sección G.3 (Histograma y polígonos) si desea que PHStat2 genere una distribución de frecuencia y un histograma.

### Polígonos de porcentaje y de porcentaje acumulado

Vea la sección G.3 (Histograma y polígonos) si desea que PHStat2 genere polígonos de porcentaje y de porcentajes acumulados como gráficas de Excel. (No hay comandos de Excel que produzcan directamente polígonos de porcentaje y de porcentaje acumulado.)

### Tablas de contingencia y gráficas de barras agrupadas

Use los Asistentes de gráficos dinámicos y de Gráficos para crear tablas de contingencia y una gráfica de barras agrupadas. Para crear una tabla de contingencia similar a la tabla 2.14 de la página 44, abra la hoja de trabajo **MUTUALFUNDS2004.XLS** en **Datos**. Seleccione **Datos** → **Informe de tablas y gráficos dinámicos** (**Datos** → **Informe de tabla dinámica** en Excel 97) y realice las siguientes entradas en la ventana de diálogo del Asistente de tablas y gráficos dinámicos:

**Paso 1:** Seleccione la opción **lista de Excel o base de datos** y la opción (si aparece) **Tabla dinámica**, y dé clic en **Siguiente**.

**Paso 2:** Ingrese **C1:J122** como **Rango** y dé clic en **Siguiente**.

**Paso 3:** Seleccione la opción **Hoja de trabajo nueva** y dé clic en el botón **Diseño**.

En la ventana de diálogo de Diseño, primero mueva una copia del rótulo **Objetivo** hacia el área de **FLA**. Después mueva una segunda copia del rótulo **Objetivo** al área de **DATOS**, lo que cambiará el rótulo a **Cuenta de objetivo**. Mueva una copia del rótulo **Riesgo** al área de **COLUMNA**. Dé clic en **Aceptar** para regresar al menú principal de la ventana de diálogo del Paso 3 y dé clic en el botón de **Opciones** para continuar.

En la ventana de diálogo de las Opciones de tabla dinámica, ingrese un nombre autodescriptivo para la tabla en el cuadro de edición **Nombre** y **0** en el cuadro de edición **Mostrar para celdas vacías**. Dé clic en **Aceptar** para regresar a la ventana de diálogo principal del paso 3 para generar la tabla dinámica.

Para crear una gráfica de barras agrupadas, dé clic en la celda fuera de la tabla bivariada, seleccione **Insertar** → **Gráfica**, y realice las siguientes entradas en las ventanas de diálogo del Asistente gráfico:

**Paso 1:** Dé clic en **Barra** en la ventana de **Gráficos de tipo estándar** y deje el primer **Subtipo de gráfico** seleccionado. Dé clic en **Siguiente**.

**Paso 2:** Con el cursor parpadeando en la ventana de **Rango de datos**, dé clic en Tabla dinámica para que Excel llene la dirección de la Tabla dinámica por usted. Dé clic en **Siguiente**.

**Paso 3:** Seleccione las opciones de formato y rótulo de gráfica para la gráfica. (Vea “Uso del Asistente para gráficos de Excel” en la página 15 para sugerencias.) Dé clic en **Siguiente**.

**Paso 4:** Seleccione **Como una hoja nueva** y dé clic en **Finalizar**. Si aparecen botones de campo en la gráfica, dé un clic derecho en cualquier botón y elija **Ocultar botones de campo gráfico dinámico** del menú corto.

*O* Vea la sección G.4 (Tablas y gráficas bivariadas) si desea que PHStat2 genere una tabla de resumen de dos variables y una gráfica agrupada.

### Diagrama de dispersión

Use el Asistente para gráficos para generar un diagrama de dispersión. Para crear un diagrama de dispersión similar al de la figura 2.15 de la página 47, abra la hoja de trabajo de **MUTUALFUNDS2004.XLS** en **Datos**. Seleccione **Insertar** → **Gráfico**, y haga las siguientes entradas en las ventanas de diálogo del Asistente gráfico:

**Paso 1:** Dé clic en **XY (Dispersión)** de la ventana de **Gráficos tipo estándar** y deje seleccionado el primer **Subtipo gráfico**. Dé clic en **Siguiente**.

**Paso 2:** Ingrese **F1:G122** en la ventana de **Rango de datos**, seleccione la opción **Columnas** y dé clic en **Siguiente**.

**Paso 3:** Seleccione las opciones de formato y rótulo de gráfica para la gráfica. (Vea “Uso del Asistente para gráficos de Excel” en la página 15 para sugerencias.) Dé clic en **Siguiente**.

**Paso 4:** Seleccione **Como una hoja nueva** y dé clic en **Finalizar**.

Esté al tanto de que el Asistente gráfico siempre supone que la primera columna de rango de datos (columna F en este ejemplo) contiene los datos de la variable **X**. Si tiene una hoja en la que los datos de la variable **Y** aparecen primero, entonces necesitará reordenar las columnas (o copiarlas en orden en una nueva hoja) antes de usar el Asistente gráfico.

## A2.2 MINITAB

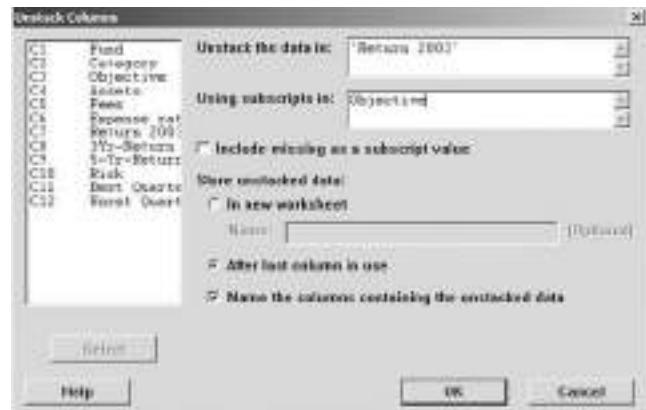
Minitab se utiliza para crear muchas de las tablas y gráficas de las que hablamos en este capítulo. Si aún no ha leído el apéndice 1.3, “Introducción al Minitab”, en la página 17, debe hacerlo ahora.

## Reclasificar datos

Los datos a menudo se ordenan para que los valores de las variables se apilen verticalmente hacia abajo en una columna. En muchos casos se requiere analizar de forma separada los diferentes subgrupos en términos de una variable numérica de interés. Por ejemplo, en los datos de los fondos de inversión, tal vez usted quiera analizar de forma separada el porcentaje del rendimiento en 2003 de los fondos de crecimiento y de los fondos de valor. Esto se logra reclasificando la variable del porcentaje de rendimiento 2003, para que, en una columna, estén localizados los porcentajes de rendimiento en 2003 de los fondos de crecimiento y, en otra, los porcentajes de rendimiento en 2003 de los fondos de valor.

Para lograr esto, abra la hoja de trabajo **MUTUALFUNDS 2004.MTW**. **Select Data → Unstack Columns** (Seleccionar Datos → Reclasificar columnas). Después haga lo siguiente:

**Paso 1:** En la ventana de diálogo con el título **Unstack Columns** (vea la figura A2.3), ingrese **C7 o Return 2003** (Rendimiento 2003) en el cuadro de edición **Unstack the data in** (Reclasificar los datos en).



**FIGURA A2.3** Ventana de diálogo para reclasificar columnas de Minitab.

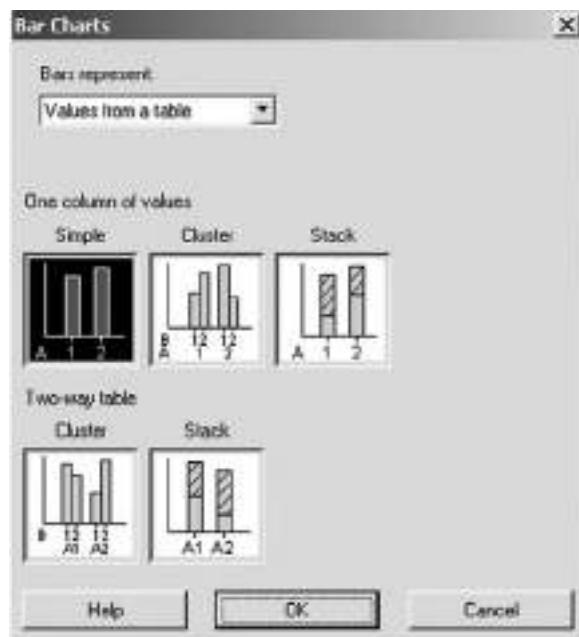
**Paso 2:** Ingrese **C3 u Objective** (Objetivo) en el cuadro de edición **Using Subscripts in:** (Usar subíndices en:)

**Paso 3:** Seleccione el botón de opción **After last column in use** (Después de la última columna en uso). Seleccione el cuadro de exploración **Name the columns containing the unstacked data** (Nombre las columnas que contienen los datos reclasificados). Dé clic en el botón **OK**. Las nuevas variables **Return2003\_Growth** y **Return2003\_Value** están ahora en las columnas **C13** y **C14**. Cambie los nombres de estas variables como lo desee.

## Gráfica de barras

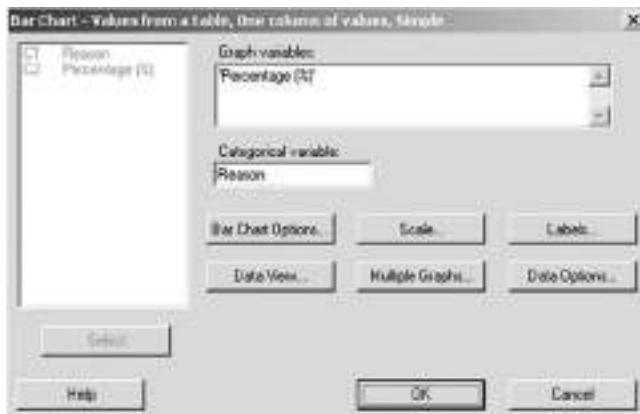
Para producir la gráfica de barras en la figura 2.1, de la página 23, abra la hoja de trabajo **ONLINESHOPPING.MTW**. Seleccione **Graph → Bar Chart** (Gráfica → Gráfica de barras) y después haga lo siguiente:

**Paso 1:** En la ventana de diálogo de **Bar charts** (vea la figura A2.4), en **Bars represent:** despliegue hacia abajo la lista, seleccione **Values from a table** (Valores de una tabla) ya que se dan las frecuencias en cada categoría. [Si está usando datos brutos como los de la hoja de trabajo **MUTUALFUNDS2004.MTW**, seleccione **Counts of unique values** (Conteo de valores únicos) en la ventana de diálogo de bars represent.] Seleccione **Simple** (Sencilla) en la ventana de gráficos. Dé clic en el botón **OK**.



**FIGURA A2.4** Ventana de diálogo gráficas de barras de Minitab.

**Paso 2:** En la ventana con el título **Values from a table**, **One column of values**, **Simple** (Valores de una tabla, una columna de valores, Ventana de diálogo sencilla), como la que aparece en la figura A2.5, ingrese **C2 o Percentage (%)** en el cuadro de edición **Graph Variables** (Variables de gráfica). Ingrese **C1 o Razón** en el cuadro de edición **Categorical Variable** (Variable categórica). Dé clic en el botón **OK**.



**FIGURA A2.5** Gráfica de barras Minitab: valores de una tabla, una columna de valores, ventana de diálogo sencilla.

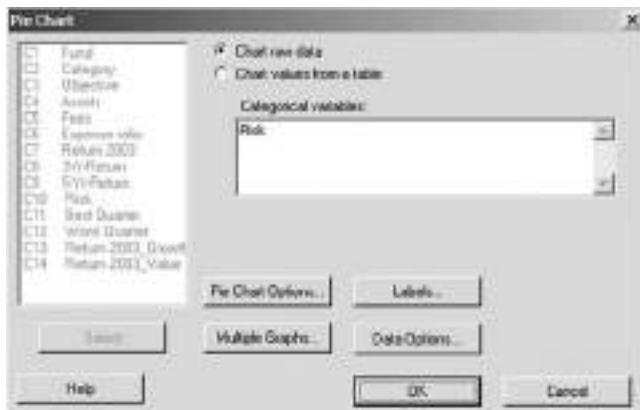
Para seleccionar los colores de las barras y los límites en la gráfica de barras:

- Paso 1:** Dé clic derecho en cualquiera de las barras de la gráfica de barras.
- Paso 2:** Seleccione **Edit bars** (Editar barras).
- Paso 3:** En la pestaña **Attributes** (Atributos) de la ventana de diálogo **Edit Bars**, ingrese las selecciones para **Fill Pattern**, **Border** y **Fill Lines** (Llenar el patrón, Límite y Llenar líneas).

### Gráfica de pastel

Para generar una gráfica de pastel similar a la de la figura 2.4 en la página 25, abra la hoja de cálculo **MUTUALFUNDS 2004.MTW**. Seleccione **Graph → Pie Chart** (Gráfica → Gráfica de pastel). Después haga lo siguiente:

- Paso 1:** En la ventana de diálogo con el título **Pie Chart** (vea la figura A2.6), seleccione la opción **Chart raw data** (Gráfica datos brutos), ya que está usando datos brutos.



**FIGURA A2.6** Ventana de diálogo para gráficas de pastel de Minitab.

tos de una hoja de trabajo. (Si utiliza las frecuencias de cada categoría como en la hoja de trabajo de **ONLINE SHOPPING.MTW**, seleccione la opción **Chart values from a table** (Valores para la gráfica de una tabla.) Ingrese **C10** o **Risk** (Riesgo) en el cuadro de edición **Categorical variables**.

- Paso 2:** Seleccione el botón **Labels** (Rótulos). En la ventana de diálogo con el título **Pie Chart—Labels** (vea la figura A2.7), seleccione la pestaña **Slice Labels** (Rótulos para rebanadas). Después seleccione las ventanas de diálogo **Category name** y **Percent** (Nombre de la categoría y Porcentaje). Dé clic en el botón **OK** para regresar a la ventana de diálogo de Gráfica de pastel. Dé clic en el botón **OK**.



**FIGURA A2.7** Ventana de diálogo gráfica de pastel —rótulos en Minitab.

### Diagrama de Pareto

Para generar el diagrama de Pareto de la figura 2.6 en la página 27, abra la hoja de trabajo **KEYBOARD.MTW**. Este conjunto de datos contiene las causas de los defectos en la columna C1 y la frecuencia de los defectos en la columna C2. Seleccione **Stat → Quality Tools → Pareto Chart** (Stat → Herramientas de Calidad → Diagrama de Pareto). En la ventana de diálogo del diagrama de Pareto (vea la figura A2.8).

- Paso 1:** Seleccione la opción **Chart defects table** (Tabla de defectos de gráfica).
- Paso 2:** En el cuadro de edición **Labels in**, ingrese **C1** o **Cause** (Causa).
- Paso 3:** En el cuadro de edición **Frequencies in**, ingrese **C2** o **Frequency** (Frecuencia).
- Paso 4:** En la ventana **Combine defects after the first** (Combinar defectos después de), ingrese **99.9**.
- Paso 5:** Dé clic en el botón **OK**.



**FIGURA A2.8** Ventana de diálogo para el diagrama de Pareto en Minitab.

Si la variable de interés estaba localizada en una única columna y está en forma bruta con cada fila indicando un tipo de error, deberá seleccionar la opción **Chart defects data in** (Datos de defectos de gráfica en), e ingresar el número de columna apropiado o el nombre de la variable en la ventana correspondiente.

### Diagrama de tallo y hojas

Para generar un diagrama de tallo y hojas del rendimiento en 2003 de todos los fondos de inversión, abra la hoja de trabajo **MUTUALFUNDS2004.MTW**. Seleccione **Graph → Stem-and-Leaf** (Gráfica → Tallo y hojas). En la ventana de diálogo **Stem-and-Leaf** (vea la figura A2.9), ingrese **C7** o **'Return 2003'** en el cuadro de edición **Graph Variables**. Dé clic en el botón **OK**.

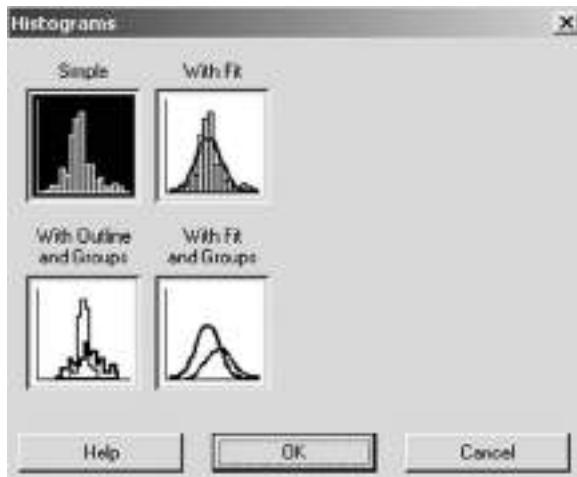


**FIGURA A2.9** Ventana de diálogo para diagrama de tallo y hojas en Minitab.

### Histograma

Para generar el histograma de los rendimientos en 2003 de todos los fondos de inversión, abra la hoja de trabajo **MUTUALFUNDS2004.MTW**. Seleccione **Graph → Histogram** (Gráfica → Histograma).

**Paso 1:** En la ventana de diálogo de **Histograms** (vea la figura A2.10) seleccione **Simple**. Dé clic en el botón **OK**.



**FIGURA A2.10** Ventana de diálogo para histogramas en Minitab.

**Paso 2:** En la ventana de diálogo **Histogram-Simple** (vea la figura A2.11), ingrese **C7** o **Return 2003** en el cuadro de edición **Graph Variables**. Dé clic en el botón **OK**.



**FIGURA A2.11** Ventana de diálogo histograma-simple de Minitab.

Para seleccionar los colores para las barras y los bordes en los histogramas,

**Paso 1:** Dé un clic derecho en cualquiera de las barras del histograma.

**Paso 2:** Seleccione **Edit bars**.

**Paso 3:** En la pestaña de **Attributes** de la ventana de diálogo **Edit Bars**, ingrese las selecciones para llenar patrones, bordes y rellenar filas.

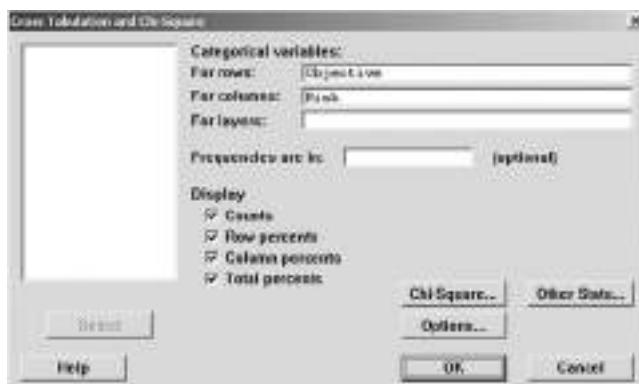
**Paso 4:** Para definir sus propios agrupamientos de clase, seleccione la pestaña **Binning**. Seleccione la opción **Midpoint** (Punto medio) para especificar los puntos medios o la opción **Cutpoints** (Puntos límite) para especificar los límites de clase. Seleccione la opción **Midpoint/Cutpoint positions**. Ingrese el conjunto de valores en el cuadro de edición.

Si desea crear histogramas separados para los fondos de crecimiento y de valor similares a los de la figura 2.9, en la página 38, primero debe reclasificar los datos (vea la página 66) y crear variables separadas para el rendimiento en el año 2003 de los fondos de crecimiento y de valor. Entonces podrá crear histogramas separados para cada uno de los dos grupos.

### Tabla de contingencia

Para generar tablas de contingencia similares a las tablas 2.14 a 2.17 de la página 44, abra la hoja de trabajo **MUTALFUNDS2004.MTW**. Seleccione **Stat → Tables → Cross Tabulation** (Tablas de contingencia) y **Chi-Square** (Chi-cuadrada).

- Paso 1:** En la ventana de diálogo Cross Tabulation y Chi-Square (vea la figura A2.12), ingrese **C3** u **Objective** en la ventana de editar For rows (para las filas). Ingrese **C10** o **Risk** en el cuadro de edición For columns (para las columnas).  
**Paso 2:** Seleccione los cuadros de exploración de **Counts**, **Row percents**, **Column percents** y **Total percents**. Dé clic en el botón **OK**.

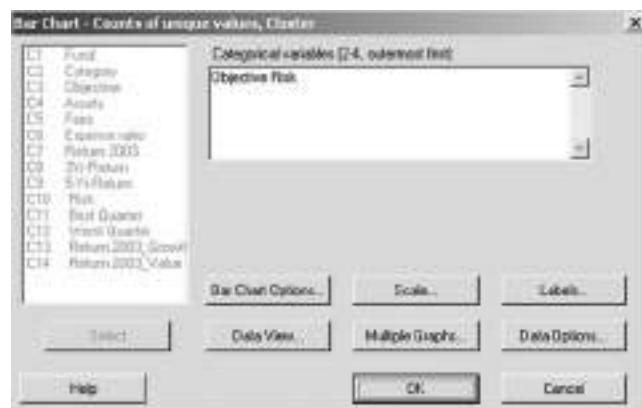


**FIGURA A2.12** Ventana de diálogo Minitab de tablas de contingencia y chi-cuadrada.

### Gráfica de barras agrupadas

Para generar una gráfica de barras agrupadas similar a la de la figura 2.14 en la página 45, abra la hoja de trabajo **MUTALFUNDS2004.MTW**. Seleccione **Graph → Bar Chart**.

- Paso 1:** En la ventana de diálogo de Bar Charts (vea la figura A2.4 en la página 66), en Bars represent, despliegue la lista hacia abajo, seleccione **Counts of unique values** puesto que está usando valores brutos. Seleccione la gráfica **Cluster** (Agrupada). Dé clic en el botón **OK**.  
**Paso 2:** En la ventana con el título Bar chart-Counts of unique values, Cluster (vea la figura A2.13), ingrese **C3** u **Objective** y **C10** o **Risk** en el cuadro de edición Categorical variables. Dé clic en el botón **OK**.

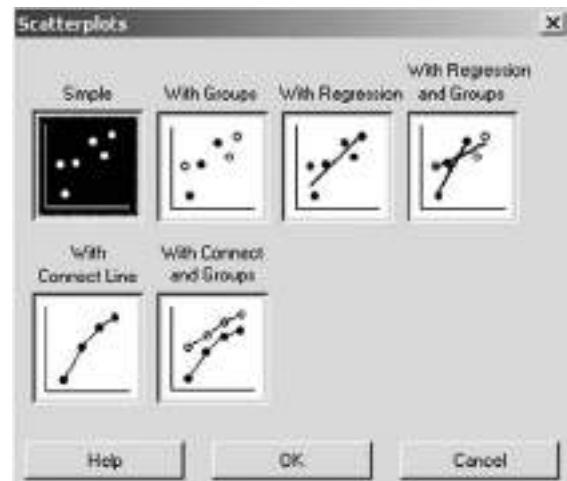


**FIGURA A2.13** Ventana de diálogo para gráficas de barra Minitab: conteo de valores únicos, agrupar.

### Diagrama de dispersión y diagrama de series de tiempo

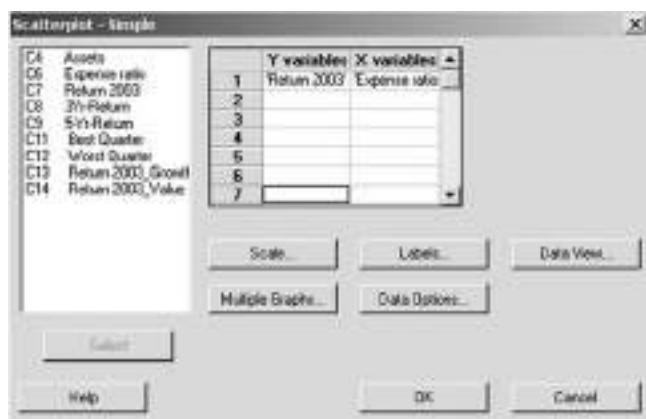
Para generar un diagrama de dispersión de la proporción de gasto del rendimiento en 2003 de los fondos de inversión (vea la figura 2.15 en la página 47), abra la hoja de trabajo **MUTALFUNDS2004.MTW**. Seleccione **Graph → Scatterplot** (Diagrama de dispersión).

- Paso 1:** En la ventana de diálogo Scatterplots (vea la figura A2.14), seleccione **Simple**. Dé clic en el botón **OK**.



**FIGURA A2.14** Ventana de diálogo para diagramas de dispersión en Minitab.

- Paso 2:** En la ventana de diálogo con el título Scatterplot-Simple (vea la figura A2.15), ingrese **C7** o **'Return 2003'** en el cuadro de edición de las variables Y en la fila 1. Ingrese **C6** o **'Expense ratio'** (Proporción de gastos) en el cuadro de edición de las variables X en la fila 1. Dé clic en el botón **OK**.



**FIGURA A2.15** Ventana de diálogo diagrama de dispersión-simple de Minitab.

Para crear un diagrama de series de tiempo, utilice el eje X para el tiempo y el eje Y para la variable de interés.

# CAPÍTULO 3

## Medidas numéricas descriptivas

---

**USO DE LA ESTADÍSTICA:** Evaluación de los rendimientos de los fondos de inversión

### 3.1 MEDIDAS DE TENDENCIA CENTRAL, VARIACIÓN Y FORMA

- La media
- La mediana
- La moda
- Cuartiles
- La media geométrica
- Rango
- Rango intercuartil
- La varianza y la desviación estándar
- Coeficiente de variación
- Puntuaciones Z
- Forma
- Exploraciones visuales: Exploración de la estadística descriptiva
- Resultado de la estadística descriptiva en Excel
- Resultado de la estadística descriptiva en Minitab

### 3.2 MEDIDAS NUMÉRICAS DESCRIPTIVAS DE UNA POBLACIÓN

- La media poblacional

Varianza y desviación estándar poblacionales

La regla empírica  
La regla de Chebyshev

### 3.3 ANÁLISIS EXPLORATORIO DE DATOS

Resumen de cinco números  
Gráfica de caja y bigote

### 3.4 LA COVARIANZA Y EL COEFICIENTE DE CORRELACIÓN

La covarianza  
Coeficiente de correlación

### 3.5 ERRORES EN LAS MEDIDAS NUMÉRICAS DESCRIPTIVAS Y CONSIDERACIONES ÉTICAS

### A.3 USO DE SOFTWARE PARA LA ESTADÍSTICA DESCRIPTIVA

- A3.1 Excel
- A3.2 Minitab
- A3.3 SPSS (*tema del CD-ROM*)

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- A describir las propiedades de tendencia central, variación y forma de los datos numéricos
- A calcular las medidas descriptivas de una población
- A construir e interpretar una gráfica de caja y bigote
- A describir la covarianza y el coeficiente de correlación

# USO DE LA ESTADÍSTICA



## Evaluación de los rendimientos de los fondos de inversión

Retomemos el estudio de los fondos de inversión presentado en el capítulo 2. Usted debe decidir en qué clases de fondos invertir. En el capítulo anterior se estudió cómo *presentar* datos en tablas y gráficas. Sin embargo, al ocuparse de datos numéricos como el rendimiento de las inversiones en los fondos de inversión durante 2003, también necesita resumir los datos y plantear preguntas estadísticas. ¿Cuál es la tendencia central del rendimiento de los diversos fondos? Por ejemplo, ¿cuál fue el rendimiento promedio de los fondos de inversión con riesgo bajo, medio y alto durante 2003? ¿Qué tanta variabilidad hay en los rendimientos? ¿El rendimiento de los fondos de alto riesgo varía más que el correspondiente a los de riesgo promedio o bajo? ¿Cómo puede utilizar esta información al decidir en cuáles fondos invertir?

Para las variables numéricas, usted necesita más que la simple imagen visual de una variable obtenida a partir de las gráficas analizadas en el capítulo 2. Por ejemplo, a usted le gustaría determinar no sólo si durante 2003 los fondos más riesgosos tuvieron un rendimiento superior, sino también si tuvieron más variación y cómo se distribuyeron en cada grupo de riesgo. También desea examinar si existe alguna relación entre el coeficiente de gastos y los rendimientos de 2003. La lectura de este capítulo le permitirá aprender sobre algunos métodos de medición:

- **Tendencia central**, es la medida que describe cómo todos los valores de los datos se agrupan en torno a un valor central.
- **Variación**, es la cantidad de dispersión o dispersión de los valores con respecto a un valor central.
- **Forma**, es el patrón de distribución de los valores desde el menor hasta el mayor.

También aprenderá sobre la covarianza y el coeficiente de correlación, que ayudan a medir la fuerza de asociación entre dos variables numéricas.

### 3.1 MEDIDAS DE TENDENCIA CENTRAL, VARIACIÓN Y FORMA

Es posible caracterizar cualquier conjunto de datos numéricos por la medición de su tendencia central, variación y forma. La mayoría de los conjuntos de datos presentan una tendencia central a agruparse en torno a un valor central. Cuando la gente habla de un “promedio”, o “valor medio”, o del valor más común o frecuente, se refiere de manera informal a la media, la mediana y la moda, tres medidas de tendencia central.

La variación mide la **distribución** o **dispersión** de valores que conforman el conjunto de datos. Una medida simple de la variación es el rango, que es la diferencia entre los valores máximo y mínimo. En la estadística, son de uso más común la desviación estándar y la varianza, dos medidas que se explican más adelante en esta sección. La forma de un conjunto de datos representa un patrón para todos los valores, desde el mínimo hasta el máximo. Como se observará más adelante en esta sección, muchos conjuntos de datos tienen un patrón semejante a una campana, cuya cima de valores está en alguna parte del centro.

## La media

La **media aritmética** (por lo general llamada la **media**) es la medida más común de la tendencia central. La media es la medida más común en la que todos los valores desempeñan el mismo papel. La media sirve como “punto de equilibrio” del conjunto de datos (como el punto de apoyo de un balancín). La media se calcula sumando todos los valores del conjunto de datos y dividiendo el resultado por el *número* de valores considerados.

Para representar a la media de una muestra, utilice el símbolo  $\bar{X}$ , llamado *X testada*. Si se considera una muestra que contiene  $n$  valores, la ecuación de su media se escribe como:

$$\bar{X} = \frac{\text{suma de los valores}}{\text{número de valores}}$$

Al utilizar la serie  $X_1, X_2, \dots, X_n$  para representar al conjunto de  $n$  valores y  $n$  para representar al número de valores, la ecuación se convierte en:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Al utilizar la notación de sumatoria (que se explica en el apéndice B), reemplace el numerador  $X_1 + X_2 + \dots + X_n$  por el término  $\sum_{i=1}^n X_i$ , que significa la suma de todos los valores  $X_i$  desde el primer valor de  $X$ , que es  $X_1$ , hasta el último valor de  $X$ , que es  $X_n$ , para formar la ecuación (3.1), una definición formal de la media de una muestra.

### MEDIA DE UNA MUESTRA

La **media de una muestra** es la suma de los valores dividida por el número de valores.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (3.1)$$

donde

$\bar{X}$  = media de la muestra

$n$  = número de valores o tamaño de la muestra

$X_i$  =  $i$ -ésimo valor de la variable  $X$

$\sum_{i=1}^n X_i$  = sumatoria de todos los valores  $X_i$  de la muestra

Como todos los valores desempeñan un papel semejante, una media se verá muy afectada por cualquier valor que difiera mucho de los demás en el conjunto de datos. Cuando tenga tales valores extremos, debe evitar el uso de la media.

La media sugiere cuál es un valor “típico” o central del conjunto de datos. Por ejemplo, si usted conoce el tiempo que le lleva arreglarse por las mañanas, podrá planear mejor su inicio del día y reducir al mínimo cualquier retraso (o adelanto) para llegar a su destino. Suponga que define en minutos (redondeando al minuto más cercano) el tiempo que le lleva arreglarse, desde que se levanta hasta que sale de casa. A lo largo de 10 días hábiles consecutivos, usted recaba los tiempos que se muestran a continuación: **TIEMPOS**

<b>Día:</b>	1	2	3	4	5	6	7	8	9	10
<b>Tiempo (minutos):</b>	39	29	43	52	39	44	40	31	44	35

El tiempo medio es 39.6 minutos, que se calculó como sigue:

$$\bar{X} = \frac{\text{suma de los valores}}{\text{número de valores}}$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\bar{X} = \frac{39 + 29 + 43 + 52 + 39 + 44 + 40 + 31 + 44 + 35}{10}$$

$$\bar{X} = \frac{396}{10} = 39.6$$

A pesar de que ni un solo día de la muestra tuvo en realidad el valor de 39.6 minutos, asignar 40 minutos a su arreglo personal sería un buen criterio para planear su inicio del día, pero sólo porque esos 10 días no contienen ningún valor extremo.

Compare lo anterior con el caso en que el valor del cuarto día fue de 102 minutos en lugar de 52. Este valor extremo provocaría que la media aumentara a 44.6 minutos, como se observa a continuación:

$$\bar{X} = \frac{\text{suma de los valores}}{\text{número de valores}}$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\bar{X} = \frac{446}{10} = 44.6$$

Un valor extremo elevó la media en más del 10%, de 39.6 a 44.6 minutos. En contraste con la media original, que estaba “en medio”, mayor que cinco de los tiempos (y menor que los otros cinco), la nueva media es mayor que 9 de los 10 tiempos de arreglo. El valor extremo provocó que la media sea una mala medida de tendencia central.

### EJEMPLO 3.1

#### EL RENDIMIENTO MEDIO EN 2003 DE LOS FONDOS DE INVERSIÓN PARA PEQUEÑOS CAPITALES

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule el rendimiento medio en 2003 de los fondos de inversión de alto riesgo para capitales reducidos.

#### SOLUCIÓN

El rendimiento medio en 2003 de los fondos de inversión para capitales reducidos (**MUTUAL-FUNDS2004**) es 51.53 calculados de la siguiente manera:

$$\bar{X} = \frac{\text{suma de los valores}}{\text{número de valores}}$$

$$= \frac{\sum_{i=1}^n X_i}{n}$$

$$= \frac{463.8}{9} = 51.53$$

El arreglo ordenado de los nueve fondos de inversión de alto riesgo para pequeños capitales es:

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

Cuatro de estos rendimientos están por debajo de la media de 51.53, y cinco están por encima de ella.

## La mediana

La **mediana** es el valor que divide en dos partes iguales a un conjunto de datos ya ordenado. La mediana no se ve afectada por los valores extremos, de manera que puede utilizarse cuando están presentes.

La mediana es el valor medio de un conjunto de datos ordenado de menor a mayor.

Para calcular la mediana del conjunto de datos, primero ordene los valores de menor a mayor. Utilice la ecuación (3.2) para calcular la clasificación del valor que corresponde a la mediana.

### MEDIANA

El 50% de los valores son menores que la mediana y el otro 50% son mayores.

$$\text{Mediana} = \frac{n+1}{2} \text{ valor clasificado} \quad (3.2)$$

Calcule el valor de la mediana siguiendo una de las dos reglas siguientes:

- **Regla 1** Si en el conjunto de datos hay un número *ímpar* de valores, la mediana es el valor colocado en medio.
- **Regla 2** Si en el conjunto de datos hay un número *par* de valores, entonces la mediana es el *promedio* de los dos valores colocados en medio.

Para calcular la mediana de la muestra de los 10 tiempos para arreglarse en las mañanas, los tiempos diarios se ordenan de la siguiente manera:

#### Valores ordenados:

29 31 35 39 39 40 43 44 44 52

#### Clasificación:

1	2	3	4	5	6	7	8	9	10
↑									
Mediana = 39.5									

Puesto que para esta muestra de 10 elementos el resultado de dividir  $n+1$  por 2 es  $(10+1)/2 = 5.5$ , debe utilizarse la regla 2 y promediar los valores clasificados quinto y sexto, 39 y 40. Por lo tanto, la mediana es 39.5. Una mediana de 39.5 significa que la mitad de los días, el tiempo necesario para arreglarse es menor o igual que 39.5 minutos, y la otra mitad de los días es mayor o igual que 39.5 minutos. Esta mediana de 39.5 minutos es muy cercana a la media del tiempo para arreglarse de 39.6 minutos.

## EJEMPLO 3.2

### CÁLCULO DE LA MEDIANA DE UNA MUESTRA CON UN NÚMERO IMPAR DE ELEMENTOS

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y con el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule la mediana del rendimiento en 2003 de los nueve fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

#### SOLUCIÓN

Puesto que para esta muestra de nueve elementos el resultado de dividir  $n+1$  por 2 es  $(9+1)/2 = 5$ , al utilizar la regla 1, la mediana es el valor clasificado como quinto. Ordene el porcentaje del rendimiento en 2003 de los nueve fondos de inversión de alto riesgo para pequeños capitales de menor a mayor:

**Valores ordenados:**

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

**Clasificación:**

1	2	3	4	5	6	7	8	9
↑								
Mediana								

La mediana del rendimiento es 53.8. La mitad de estos fondos de inversión tienen rendimientos iguales o menores que 53.8 y la otra mitad tiene rendimientos iguales o superiores.

**La moda**

La **moda** es el valor del conjunto de datos que aparece con mayor frecuencia. Al igual que en la mediana y a diferencia de la media, los valores extremos no afectan a la moda. Usted sólo debe utilizar la media con propósitos descriptivos, ya que varía más de una muestra a otra que la media o la mediana. Con frecuencia, en un conjunto de datos no existe moda, o bien, hay varias modas. Por ejemplo, considere los datos de tiempo para arreglarse que se muestran a continuación.

29 31 35 39 39 40 43 44 44 52

Existen dos modas, 39 y 44 minutos, ya que cada uno de estos valores aparece dos veces.

**EJEMPLO 3.3****CÁLCULO DE LA MODA**

El gerente de sistemas encargado de la red de una empresa lleva un registro del número de fallas del servidor que se presentan por día. Calcule la moda de los siguientes datos, que representan el número de fallas diarias del servidor durante las últimas dos semanas.

1 3 0 3 26 2 7 4 0 2 3 3 6 3

**SOLUCIÓN**

El arreglo ordenado de estos datos es:

0 0 1 2 2 3 3 3 3 3 4 6 7 26

Como el 3 aparece cinco veces, más que ningún otro valor, la moda es 3. De esta forma, el gerente de sistemas se dará cuenta de que la situación más común es la presencia de tres fallas del servidor al día. Para este conjunto de datos, la mediana también es igual a 3, mientras que la media es de 4.5. El valor extremo de 26 es atípico. Con estos datos, la mediana y la moda miden la tendencia central mejor que la media.

Un conjunto de datos no tiene moda cuando ninguno de los valores es “más frecuente”. En el ejemplo 3.4 aparece un conjunto de datos sin moda.

**EJEMPLO 3.4****DATOS SIN MODA**

Calcule la moda del rendimiento medio en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

**SOLUCIÓN**

El arreglo ordenado para estos datos es:

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

Estos datos no tienen moda. Ninguno de sus valores aparece con mayor frecuencia; cada uno aparece sólo una vez.

## Cuartiles

Los **cuartiles** dividen a un conjunto de datos en cuatro partes iguales: el **primer cuartil  $Q_1$**  separa al 25.0%, que abarca a los valores más pequeños, del 75.0% restante, constituido por los que son mayores. El **segundo cuartil  $Q_2$**  es la mediana: 50.0% de sus valores son menores que la mediana y 50.0% son mayores. El **tercer cuartil  $Q_3$**  separa al 75.0%, que abarca a los valores más grandes, del 25.0% restante constituido por los que son menores. Las ecuaciones (3.3) y (3.4) definen a los cuartiles primero y tercero.<sup>1</sup>

<sup>1</sup>El  $Q_1$ , la mediana y el  $Q_3$  también son el 25, 50 y 75º percentil, respectivamente. Por lo general, las ecuaciones (3.2), (3.3) y (3.4) se expresan en términos de cálculo de percentiles: percentil ( $p * 100$ )º = valor clasificado  $p * (n + 1)$ .

### PRIMER CUARTIL $Q_1$

El 25.0% de los valores son menores que el primer cuartil  $Q_1$ , y el 75.0% son mayores que el primer cuartil  $Q_1$ .

$$Q_1 = \frac{n + 1}{4} \text{ valor clasificado} \quad (3.3)$$

### TERCER CUARTIL $Q_3$

El 75.0% de los valores son menores que el tercer cuartil  $Q_3$ , y el 25.0% son mayores que el tercer cuartil  $Q_3$ .

$$Q_3 = \frac{3(n + 1)}{4} \text{ valor clasificado} \quad (3.4)$$

Para calcular los cuartiles, se utilizan las siguientes reglas:

- **Regla 1** Si el resultado es un número entero, entonces el cuartil es igual al valor clasificado. Por ejemplo, si el tamaño de la muestra es  $n = 7$ , el primer cuartil  $Q_1$  es igual a  $(7 + 1)/4 =$  segundo valor clasificado.
- **Regla 2** Si el resultado es una fracción de mitad (2.5, 4.5, etcétera), entonces el cuartil es igual al promedio de los valores clasificados correspondientes. Por ejemplo, si el tamaño de la muestra es  $n = 9$ , el primer cuartil  $Q_1$  es igual al valor clasificado como  $(9 + 1)/4 = 2.5$ , la mitad entre los valores clasificados como segundo y tercero.
- **Regla 3** Si el resultado no es un número entero ni una fracción de mitad, se redondea al entero más cercano y se selecciona ese valor clasificado. Por ejemplo, si el tamaño de la muestra es  $n = 10$ , el primer cuartil  $Q_1$  es igual a  $(10 + 1)/4 =$  valor clasificado como 2.75. Se redondea el 2.75 a 3 y se utiliza en valor clasificado como tercero.

Con el fin de ilustrar el cálculo de los cuartiles para los datos referentes a los tiempos para arreglarse, se ordenan de menor a mayor.

#### Valores ordenados:

29 31 35 39 39 40 43 44 44 52

#### Clasificación:

1 2 3 4 5 6 7 8 9 10

El primer cuartil es el valor clasificado como  $(n + 1)/4 = (10 + 1)/4 = 2.75$ . Al emplear la tercera regla de los cuartiles, redondeamos al tercer valor clasificado. Para los datos sobre el tiempo necesario para arreglarse, el valor clasificado como tercero es 35 minutos. Interprete el primer cuartil de 35 como que el 25% de los días el tiempo necesario para arreglarse es menor o igual a 35 minutos, y que el 75% de los días ese tiempo es mayor o igual a 35 minutos.

El tercer cuartil es el valor clasificado como  $3(n + 1)/4 = 3(10 + 1)/4 = 8.25$ . Empleando la tercera regla de los cuartiles, redondeamos al valor clasificado como octavo. El valor clasificado como octavo en los datos del tiempo necesario para arreglarse es de 44 minutos. Interprete esto como que el 75% de los días, el tiempo necesario para arreglarse es menor o igual que 44 minutos, y que el 25% de los días ese tiempo es mayor o igual que 44 minutos.

**EJEMPLO 3.5****CÁLCULO DE LOS CUARTILES**

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño de capital invertido (pequeño, mediano y gran capital). Calcule el primer cuartil ( $Q_1$ ) y el tercer cuartil ( $Q_3$ ) del rendimiento en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. **MUTUAL-FUNDS2004**

**SOLUCIÓN**

Ordenados de menor a mayor, los porcentajes de rendimiento de los nueve fondos de inversión de alto riesgo para pequeños capitales durante 2003 son:

**Valor clasificado:**

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

**Clasificación:**

1 2 3 4 5 6 7 8 9

Para estos datos:

$$\begin{aligned} Q_1 &= \frac{(n+1)}{4} \text{ valor clasificado} \\ &= \frac{9+1}{4} = 2.5 \text{ valor clasificado} \end{aligned}$$

Por lo tanto, al utilizar la segunda regla, resulta que  $Q_1$  es el valor clasificado como 2.5, que está justo a la mitad entre los valores clasificados como segundo y tercero. Como el valor clasificado como segundo es 39.2 y el tercero es 44.2, el primer cuartil  $Q_1$  es el que está justo en medio de 39.2 y 44.2. De esta forma,

$$Q_1 = \frac{39.2 + 44.2}{2} = 41.7$$

Para encontrar el tercer cuartil  $Q_3$ :

$$\begin{aligned} Q_3 &= \frac{3(n+1)}{4} \text{ valor clasificado} \\ &= \frac{3(9+1)}{4} = 7.5 \text{ valor clasificado} \end{aligned}$$

Así, al utilizar la segunda regla,  $Q_3$  es el valor clasificado entre los valores séptimo y octavo. Como el valor clasificado como séptimo es 59.3 y el octavo es 62.4, el tercer cuartil  $Q_3$  es el que está justo en medio de 59.3 y 62.4. De esta forma,

$$Q_3 = \frac{59.3 + 62.4}{2} = 60.85$$

Un primer cuartil de 41.7 señala que el 25% de los rendimientos obtenidos durante 2003 por los fondos de alto riesgo para pequeños capitales fueron menores o iguales que 41.7, mientras que el 75% de ellos fueron mayores o iguales que 41.7. El tercer cuartil de 60.85 indica que el 75% de los rendimientos obtenidos durante el mismo año por los fondos de alto riesgo para pequeños capitales fueron menores o iguales que 60.85 y que el 25% fueron mayores o iguales que 60.85.

## La media geométrica

La media geométrica y la razón geométrica de rendimiento miden el estado de una inversión en el tiempo. La **media geométrica** mide la razón de cambio de una variable en el tiempo. La ecuación 3.5 define a la media geométrica.

### MEDIA GEOMÉTRICA

La media geométrica es la raíz  $n$ -ésima del producto de  $n$  valores

$$\bar{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n} \quad (3.5)$$

La ecuación 3.6 define a la media geométrica de la tasa de rendimiento.

### MEDIA GEOMÉTRICA DE LA TASA DE RENDIMIENTO

$$\bar{R}_G = [(1 + R_1) \times (1 + R_2) \times \cdots \times (1 + R_n)]^{1/n} - 1 \quad (3.6)$$

donde

$R_i$  es la tasa de rendimiento durante el periodo  $i$

Para ilustrar el uso de estas medidas, considere una inversión de \$100,000 que se reduce hasta tener un valor de \$50,000 al final del año 1 y luego recupera su valor original de \$100,000 al finalizar el año 2. La tasa de rendimiento de esta inversión en el periodo de dos años es 0, porque los valores inicial y final permanecen sin cambio. Sin embargo, la media aritmética de las tasas de rendimiento anuales de esta inversión es

$$\bar{X} = \frac{(-0.50) + (1.00)}{2} = 0.25 \text{ o } 25\%$$

ya que la tasa de rendimiento del año 1 es

$$R_1 = \left( \frac{50,000 - 100,000}{100,000} \right) = -0.50 \text{ o } -50\%$$

y la tasa de rendimiento del año 2 es

$$R_2 = \left( \frac{100,000 - 50,000}{50,000} \right) = 1.00 \text{ o } 100\%$$

Al utilizar la ecuación (3.6), se sabe que la media geométrica de la tasa de rendimiento para los dos años es

$$\begin{aligned} \bar{R}_G &= [(1 + R_1) \times (1 + R_2)]^{1/2} - 1 \\ &= [(1 + (-0.50)) \times (1 + (1.00))]^{1/2} - 1 \\ &= [(0.50) \times (2.0)]^{1/2} - 1 \\ &= [1.0]^{1/2} - 1 \\ &= 1 - 1 = 0 \end{aligned}$$

Por lo tanto, la media geométrica de la tasa de rendimiento refleja con mayor exactitud el cambio (cero) del valor de la inversión durante el periodo de dos años de la media aritmética.

**EJEMPLO 3.6****CALCULE LA MEDIA GEOMÉTRICA DE LA TASA DE RENDIMIENTO**

El porcentaje de cambio del índice compuesto NASDAQ fue del  $-31.53\%$  en 2002 y del  $+50.01\%$  en 2003. Calcule la tasa geométrica de rendimiento.

**SOLUCIÓN**

Al utilizar la ecuación (3.6), se sabe que la media geométrica de la tasa de rendimiento del índice NASDAQ para los dos años es

$$\begin{aligned}\bar{R}_G &= [(1 + R_1) \times (1 + R_2)]^{1/n} - 1 \\ &= [(1 + (-0.3153)) \times (1 + (0.5001))]^{1/2} - 1 \\ &= [(0.6847) \times (1.5001)]^{1/2} - 1 \\ &= [1.0271]^{1/2} - 1 \\ &= 1.0135 - 1 = 0.0135\end{aligned}$$

La media geométrica de la tasa de rendimiento del índice NASDAQ para los dos años es del  $1.35\%$ .

**Rango**

El **rango** es la medida numérica descriptiva más sencilla de la variación en un conjunto de datos.

**RANGO**

El rango es igual al valor mayor menos el valor menor.

$$\text{Rango} = X_{\text{mayor}} - X_{\text{menor}} \quad (3.7)$$

Para determinar el rango de los tiempos necesarios para arreglarse, los datos se ordenan de menor a mayor:

29 31 35 39 39 40 43 44 44 52

Al emplear la ecuación (3.7), se sabe que el rango es de  $52 - 29 = 23$  minutos. Un rango de 23 minutos señala que la mayor diferencia del tiempo necesario para arreglarse por la mañana entre dos días cualesquiera es de 23 minutos.

**EJEMPLO 3.7****CALCULE EL RANGO DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE ALTO RIESGO PARA PEQUEÑOS CAPITALES**

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule el rango del rendimiento en 2003 de los nueve fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

**SOLUCIÓN**

Ordenados de menor a mayor, los rendimientos en 2003 de los nueve fondos de inversión de alto riesgo para pequeños capitales son:

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

Por lo tanto, al utilizar la ecuación 3.7, se sabe que el rango =  $66.5 - 37.3 = 29.2$ .

La mayor diferencia entre dos rendimientos cualesquiera de los fondos de inversión de alto riesgo para pequeños capitales es de 29.2.

El rango mide la *distribución total* del conjunto de datos. Aunque el rango es una medida simple de la variación total de los datos, no toma en cuenta *cómo* se distribuyen los datos entre los valores menor y mayor. En otras palabras, el rango no indica si los valores están distribuidos de manera uniforme a todo lo largo del conjunto de datos, agrupados cerca de la parte media, o agrupados cerca de uno o ambos extremos. De esta manera, resulta engañoso utilizar el rango como medida de la variación cuando al menos uno de los valores es extremo.

### Rango intercuartil

El **rango intercuartil** (también llamado **dispersión media**) es la diferencia entre el *tercer y primer cuartil* de un conjunto de datos.

#### RANGO INTERCUARTIL

El rango intercuartil es la diferencia entre los cuartiles tercero y primero.

$$\text{Rango intercuartil} = Q_3 - Q_1 \quad (3.8)$$

El rango intercuartil mide la dispersión en la mitad (parte central) de los datos, así que no se ve influido por los valores extremos. Para determinar el rango intercuartil de los tiempos necesarios para arreglarse

29 31 35 39 39 40 43 44 44 52

utilice la ecuación (3.8) y los resultados obtenidos en la página 77,  $Q_1 = 35$  y  $Q_3 = 44$ .

$$\text{Rango intercuartil} = 44 - 35 = 9 \text{ minutos}$$

Por lo tanto, el rango intercuartil del tiempo necesario para arreglarse es de 9 minutos. Por lo general, al intervalo de 35 a 44 se le denomina *la mitad media*.

### EJEMPLO 3.8

#### CALCULE EL RANGO INTERCUARTIL DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE ALTO RIESGO PARA PEQUEÑOS CAPITALES

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule el rango intercuartil del rendimiento en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

#### SOLUCIÓN

Ordenados de menor a mayor, los rendimientos de los nueve fondos de inversión de alto riesgo para pequeños capitales durante 2003 son:

37.3 39.2 44.2 44.5 53.8 56.6 59.3 62.4 66.5

Utilice la ecuación 3.8 y los resultados obtenidos en la página 78,  $Q_1 = 41.7$  y  $Q_3 = 60.85$ .

$$\text{Rango intercuartil} = 60.85 - 41.7 = 19.15$$

Así, el rango intercuartil de los rendimientos en 2003 es de 19.15.

Como el rango intercuartil no toma en cuenta ningún valor menor que  $Q_1$  ni mayor que  $Q_3$ , no se ve afectado por los valores extremos. Las medidas de resumen como la mediana,  $Q_1$ ,  $Q_3$ , y el rango intercuartil, que no reciben la influencia de valores extremos, se denominan medidas resistentes.

## La varianza y la desviación estándar

A pesar de que el rango y el rango intercuartil son medidas de la variación, no contemplan *cómo* se distribuyen o se agrupan los valores que están entre los extremos. La **varianza** y la **desviación estándar** son dos medidas de la variación muy utilizadas para tomar en cuenta cómo se distribuyen los datos. Estos estadísticos miden la dispersión “promedio” alrededor de la media, es decir, qué tanto varían los valores más grandes que están por encima de ella y cómo se distribuyen los valores menores que están por debajo de ella.

Una medida simple de la variación alrededor de la media consideraría la diferencia entre cada uno de los valores y la media, y luego las sumaría. Sin embargo, si usted hiciera eso, podría descubrir que la media es el punto de equilibrio de un conjunto de datos y que tales diferencias sumarían cero en *todo* conjunto de datos. Una medida de la variación que sería distinta de un conjunto de datos a otro consistiría en *elevar al cuadrado* la diferencia entre cada uno de los valores y la media, y después sumarlas. En estadística, esta cantidad se denomina **suma de cuadrados** (o  $SS$ ). Esta suma luego se divide entre el número de valores menos 1 (para datos de la muestra), con el fin de obtener una varianza de la muestra ( $S^2$ ). La raíz cuadrada de la varianza de la muestra es la desviación estándar de la muestra ( $S$ ).

Puesto que la suma de cuadrados es una suma de diferencias elevadas al cuadrado que, por las reglas aritméticas siempre será no negativa, *ni la varianza ni la desviación estándar podrán ser negativas*. En casi todos los conjuntos de datos, la varianza y la desviación estándar tendrán un valor positivo, aunque si no existe variación en todo el conjunto de datos y todos los valores de la muestra son los mismos, ambos estadísticos serán igual a cero.

En una muestra que contiene  $n$  valores,  $X_1, X_2, X_3, \dots, X_n$ , la varianza de la muestra (representada por el símbolo  $S^2$ ) es

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}$$

La ecuación 3.9 expresa esta ecuación utilizando la notación de sumatoria.

### VARIANZA PARA UNA MUESTRA

La **varianza para una muestra** es la suma de las diferencias con respecto a la media elevada al cuadrado y dividida por el tamaño de la muestra menos uno.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \quad (3.9)$$

donde

$\bar{X}$  = media

$n$  = tamaño de la muestra

$X_i$  =  $i$ -ésimo valor de la variable  $X$

$\sum_{i=1}^n (X_i - \bar{X})^2$  = sumatoria de los cuadrados de todas las diferencias entre los valores de  $X_i$  y  $\bar{X}$ .

### DESVIACIÓN ESTÁNDAR DE LA MUESTRA

La **desviación estándar de una muestra** es la raíz cuadrada de la suma de los cuadrados de las diferencias con respecto a la media dividida por el tamaño de la muestra menos uno.

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \quad (3.10)$$

Si el denominador fuese  $n$  en vez de  $n - 1$ , la ecuación (3.9) [y el término interno de la ecuación (3.10)] calcularía el promedio de las diferencias con respecto a la media elevadas al cuadrado. Sin embargo, se utiliza  $n - 1$  porque ciertas propiedades matemáticas convenientes del estadístico  $S^2$  lo hacen apropiado para la inferencia estadística (que analizaremos en el capítulo 7). Conforme aumenta el tamaño de la muestra, se hace cada vez más pequeña la diferencia entre dividir por  $n$  o por  $n - 1$ .

Es más probable que usted utilice la desviación estándar de la muestra como medida de la variación [definida en la ecuación (3.10)]. A diferencia de la varianza de la muestra, que es una cantidad elevada al cuadrado, la desviación estándar siempre es un número con las mismas unidades que los datos de muestra originales. La desviación estándar le ayuda a conocer de qué manera se agrupan o distribuyen un conjunto de datos con respecto a su media. En casi todos los conjuntos de datos, la mayoría de los valores observados quedan dentro de un intervalo de más menos una desviación estándar por encima y por debajo de la media. Por esa razón, conocer la media y la desviación estándar ayuda a definir por lo menos dónde se agrupa la mayoría de los valores de los datos.

Para calcular a mano la varianza  $S^2$  y la desviación estándar  $S$  de una muestra:

**Paso 1:** Calcule la diferencia entre cada uno de los valores y la media.

**Paso 2:** Eleve al cuadrado cada una de esas diferencias.

**Paso 3:** Sume las diferencias elevadas al cuadrado.

**Paso 4:** Divida el total entre  $n - 1$ , para obtener la varianza de la muestra.

**Paso 5:** Extraiga la raíz cuadrada de la varianza de la muestra, para obtener la desviación estándar de la muestra.

La tabla 3.1 muestra los cuatro primeros pasos para calcular la varianza de los datos referentes al tiempo necesario para arreglarse, con una media ( $\bar{X}$ ) = 39.6 (vea el cálculo de la media en la página 74). En la segunda columna se muestra el paso 1. En la tercera columna se muestra el paso 2. En la parte inferior se muestra la suma de las diferencias elevadas al cuadrado (paso 3). Luego, este total se divide entre  $10 - 1 = 9$ , para calcular la varianza (paso 4).

**TABLA 3.1**

Cálculo de la varianza del tiempo necesario para arreglarse.

$\bar{X} = 39.6$		
<b>Tiempo</b> <b>(<math>X</math>)</b>	<b>Paso 1:</b> <b><math>(X_i - \bar{X})</math></b>	<b>Paso 2:</b> <b><math>(X_i - \bar{X})^2</math></b>
39	-0.60	0.36
29	-10.60	112.36
43	3.40	11.56
52	12.40	153.76
39	-0.60	0.36
44	4.40	19.36
40	0.40	0.16
31	-8.60	73.96
44	4.40	19.36
35	-4.60	21.16
<b>Paso 3:</b> <b>Suma:</b>		<b>Paso 4:</b> <b>Dividido por (<math>n - 1</math>):</b>
412.40		45.82

También es posible calcular la varianza si se sustituyen los valores de los términos en la ecuación 3.9:

$$\begin{aligned}
 S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\
 &= \frac{(39 - 39.6)^2 + (29 - 39.6)^2 + \dots + (35 - 39.6)^2}{10 - 1} \\
 &= \frac{412.4}{9} \\
 &= 45.82
 \end{aligned}$$

Puesto que la varianza está en unidades cuadradas (en minutos cuadrados en este caso), para calcular la desviación estándar se calcula la raíz cuadrada de la varianza. Al utilizar la ecuación (3.10) de la página 82, la desviación estándar  $S$  de la muestra es:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{45.82} = 6.77$$

Esto indica que los tiempos necesarios para arreglarse en esta muestra se agrupan dentro de los 6.77 minutos que circundan a la media de 39.6 minutos (es decir, se agrupan entre  $\bar{X} - 1S = 32.83$  y  $\bar{X} + 1S = 46.37$ ). De hecho, 7 de los 10 quedan dentro de este intervalo.

Al utilizar la segunda columna de la tabla 3.1, también es posible calcular que la suma de las diferencias entre cada uno de los valores y la media es cero. Para todo conjunto de datos, esta suma siempre será igual a cero:

$$\sum_{i=1}^n (X_i - \bar{X}) = 0 \text{ para todos los conjuntos de datos}$$

Esta propiedad es una de las razones por las que la media se utiliza como la medida más común de tendencia central.

### EJEMPLO 3.9

#### CÁLCULO DE LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN PARA PEQUEÑOS CAPITALES

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule la varianza y la desviación estándar del rendimiento en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

#### SOLUCIÓN

La tabla 3.2 ilustra el cálculo de la varianza y la desviación estándar del rendimiento en 2003 para los fondos de inversión de alto riesgo para pequeños capitales. Utilice la ecuación (3.9) de la página 82:

$$\begin{aligned}
 S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\
 &= \frac{(44.5 - 51.53)^2 + (39.2 - 51.53)^2 + \dots + (66.5 - 51.53)^2}{9 - 1} \\
 &= \frac{891.16}{8} \\
 &= 111.395
 \end{aligned}$$

**TABLA 3.2**

Cálculo de la varianza del rendimiento en 2003 para los fondos de inversión de alto riesgo para pequeños capitales.

$\bar{X} = 51.5333$			
<b>Rendimiento</b>	<b>Paso 1:</b> $(X_i - \bar{X})$	<b>Paso 2:</b> $(X_i - \bar{X})^2$	
<b>2003</b>			
44.5	-7.0333	49.4678	
39.2	-12.3333	152.1111	
62.4	10.8667	118.0844	
59.3	7.7667	60.3211	
56.6	5.0667	25.6711	
53.8	2.2667	5.1378	
37.3	-14.2333	202.5878	
44.2	-7.3333	53.7778	
66.5	14.9667	224.0011	
		<b>Paso 3:</b> <b>Suma:</b>	<b>Paso 4:</b> <b>Dividido por <math>(n - 1)</math>:</b>
		891.16	111.395

Al utilizar la ecuación (3.10) de la página 82, se sabe que la desviación estándar  $S$  de la muestra es:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{111.395} = 10.55$$

La desviación estándar de 10.55 indica que los rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales se agrupan dentro de los 10.55 que rodean a la media de 51.53 (es decir, se agrupan entre  $\bar{X} - 1S = 40.98$  y  $\bar{X} + 1S = 62.08$ ). De hecho, el 55.6% (5 de 9) de los rendimientos en 2003 quedan dentro de este intervalo.

A continuación se resumen las características del rango, del rango intercuartil, de la varianza y de la desviación estándar.

- Cuanto más esparcidos o dispersos están los datos, son mayores el rango, el rango intercuartil, la varianza y la desviación estándar.
- Cuanto más concentrados u homogéneos son los datos, son menores el rango, el rango intercuartil, la varianza y la desviación estándar.
- Si todos los valores son los mismos (de tal manera que no hay variación de los datos), el rango, el rango intercuartil, la varianza y la desviación estándar son iguales a cero.
- Ninguna de las medidas de la variación (rango, rango intercuartil, desviación estándar y varianza) puede ser negativa.

## Coeficiente de variación

A diferencia de las medidas de la variación antes expuestas, el **coeficiente de variación** es una medida relativa de la variación que siempre se expresa como porcentaje, más que en términos de las unidades de los datos en particular. El coeficiente de variación, que se denota mediante el símbolo  $CV$ , mide de dispersión de los datos con respecto a la media.

## COEFICIENTE DE VARIACIÓN

El coeficiente de variación es igual a la desviación estándar dividida por la media, multiplicada por 100%.

$$CV = \left( \frac{S}{\bar{X}} \right) 100\% \quad (3.11)$$

donde

$S$  = desviación estándar de la muestra

$\bar{X}$  = media de la muestra

Para la muestra de los 10 tiempos para arreglarse, como  $\bar{X} = 39.6$  y  $S = 6.77$ , el coeficiente de variación es

$$CV = \left( \frac{S}{\bar{X}} \right) 100\% = \left( \frac{6.77}{39.6} \right) 100\% = 17.10\%$$

Para estos datos, la desviación estándar es el 17.1% del tamaño de la media.

El coeficiente de variación es muy útil al comparar dos o más conjuntos de datos medidos con unidades distintas, como ilustra el ejemplo 3.10.

**EJEMPLO 3.10**

## COMPARACIÓN DE DOS COEFICIENTES DE VARIACIÓN CUANDO DOS VARIABLES TIENEN DISTINTAS UNIDADES DE MEDIDA

El gerente de operaciones de un servicio de entrega de paquetería está pensando si es conveniente adquirir una nueva flota de camiones. Al guardar los paquetes en los camiones para su entrega, se deben tomar en cuenta dos características principales: el peso (en libras) y el volumen (en pies cúbicos) de cada artículo.

El gerente de operaciones toma una muestra de 200 paquetes, y encuentra que la media del peso es 26.0 libras, con una desviación estándar de 3.9 libras, mientras que la media en volumen es de 8.8 pies cúbicos, con una desviación estándar de 2.2 pies cúbicos. ¿Cómo puede el gerente de operaciones comparar la variación de peso y volumen?

**SOLUCIÓN**

Como las unidades difieren para el peso y volumen, el gerente de operaciones debe comparar la variabilidad relativa en ambos tipos de medidas.

Para el peso, el coeficiente de variación es

$$CV_W = \left( \frac{3.9}{26.0} \right) 100\% = 15.0\%$$

para el volumen, el coeficiente de variación es

$$CV_V = \left( \frac{2.2}{8.8} \right) 100\% = 25.0\%$$

De esta forma, en relación con la media el volumen del paquete es mucho más variable que su peso.

**Puntuaciones Z**

Un **valor extremo** o **atípico** es un valor ubicado muy lejos de la media. Las puntuaciones Z son útiles para identificar atípicos. Cuanto mayor es la puntuación Z, mayor es la distancia entre tal valor y la media. La **puntuación Z** es igual a la diferencia entre ese valor y la media, dividida por la desviación estándar.

## PUNTUACIONES Z

$$Z = \frac{X - \bar{X}}{S} \quad (3.12)$$

Si se consideran los tiempos necesarios para arreglarse por la mañana, se observa que la media es de 39.6 minutos y la desviación estándar de 6.77 minutos. El tiempo necesario para arreglarse el primer día es de 39.0 minutos. La puntuación Z para el día 1 se calcula a partir de

$$\begin{aligned} Z &= \frac{X - \bar{X}}{S} \\ &= \frac{39.0 - 39.6}{6.77} \\ &= -0.09 \end{aligned}$$

La tabla 3.3 muestra las puntuaciones Z de los 10 días. La mayor es de 1.83 para el día 4, cuando el tiempo necesario para arreglarse fue de 52 minutos. La menor fue -1.57 para el día 2, cuando el tiempo necesario para arreglarse fue de 29 minutos. Como regla general, una puntuación Z se considera atípica si es menor que -3.0 o mayor que +3.0. Ninguno de los tiempos satisface este criterio.

**TABLA 3.3**

Puntuaciones Z para los 10 tiempos necesarios para arreglarse.

	Tiempo (X)	Puntuación Z
	39	-0.09
	29	-1.57
	43	0.50
	52	1.83
	39	-0.09
	44	0.65
	40	0.06
	31	-1.27
	44	0.65
	35	-0.68
<b>Media</b>	39.6	
<b>Desviación estándar</b>	6.77	

**EJEMPLO 3.11****CÁLCULO DE LAS PUNTUACIONES Z DEL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE ALTO RIESGO PARA PEQUEÑOS CAPITALES**

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule las puntuaciones Z del rendimiento en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

**SOLUCIÓN**

La tabla 3.4 ilustra las puntuaciones Z de los rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. La puntuación Z más grande es 1.42, correspondiente a un rendimiento porcentual de 66.5. La puntuación Z más baja es -1.35, correspondiente a un rendimiento porcentual de 37.3. Como regla general, se considera que una puntuación Z es atípica si es menor que -3.0 o mayor que +3.0. Ninguno de los rendimientos porcentuales satisface el criterio para considerarlo atípico.

**TABLA 3.4**

Puntuación Z del rendimiento en 2003 de los fondos de inversión de alto riesgo para pequeños capitales.

	Rendimiento 2003	Puntuaciones Z
	44.5	-0.67
	39.2	-1.17
	62.4	1.03
	59.3	0.74
	56.6	0.48
	53.8	0.21
	37.3	-1.35
	44.2	-0.69
	66.5	1.42
<b>Media</b>	51.53	
<b>Desviación estandar</b>	10.55	

## Forma

Una tercera e importante propiedad que describe a un conjunto de datos numéricos es la forma. Forma es el patrón de distribución de los valores de los datos a través del rango de todos los valores. La distribución puede ser **simétrica** cuando los valores pequeños y grandes se equilibrان entre sí, o **asimétrica**, cuando muestra desequilibrio de los valores pequeños o grandes.

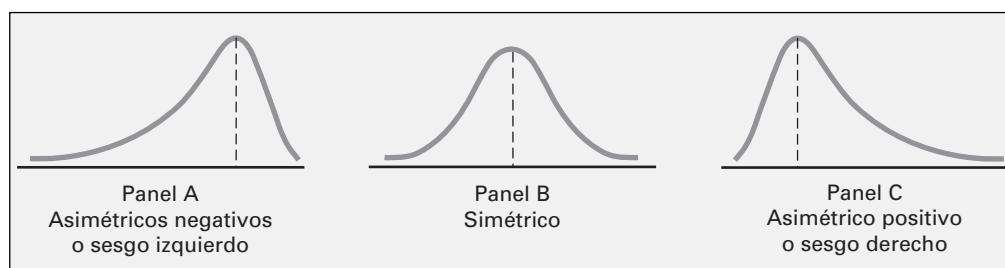
La forma influye en la relación de la media con la mediana de las siguientes maneras:

- Media < mediana; asimétrica negativa o sesgo izquierdo.
- Media = mediana; simétrica o asimetría cero.
- Media > mediana; asimétrica positiva o sesgo derecho.

La figura 3.1 describe tres conjuntos de datos, cada uno con distinta forma.

**FIGURA 3.1**

Comparación de tres conjuntos de datos con distinta forma.



Los datos del panel A son negativos, o **sesgados a la izquierda**. En este panel, la mayoría de los valores están en la parte superior de la distribución. Existe una cola larga y la distorsión hacia la izquierda es provocada por algunos valores muy pequeños. Estos valores extremadamente pequeños empujan la media hacia abajo, de manera que la media es menor que la mediana.

Los datos del panel B son simétricos. Cada mitad de la curva es una imagen al espejo del otro. Los valores bajos y altos de la escala se equilibrان, y la media es igual a la mediana.

Los datos del panel C son **asimétricos positivos o sesgados a la derecha**. En este panel, la mayoría de los valores están en la parte inferior de la distribución. Existe una larga cola a la derecha de la distribución y cierta distorsión hacia la derecha provocada por algunos valores muy grandes. Estos valores sumamente grandes empujan a la media hacia arriba, de manera que la media resulta mayor que la mediana.

## Resultados de la estadística descriptiva en Excel

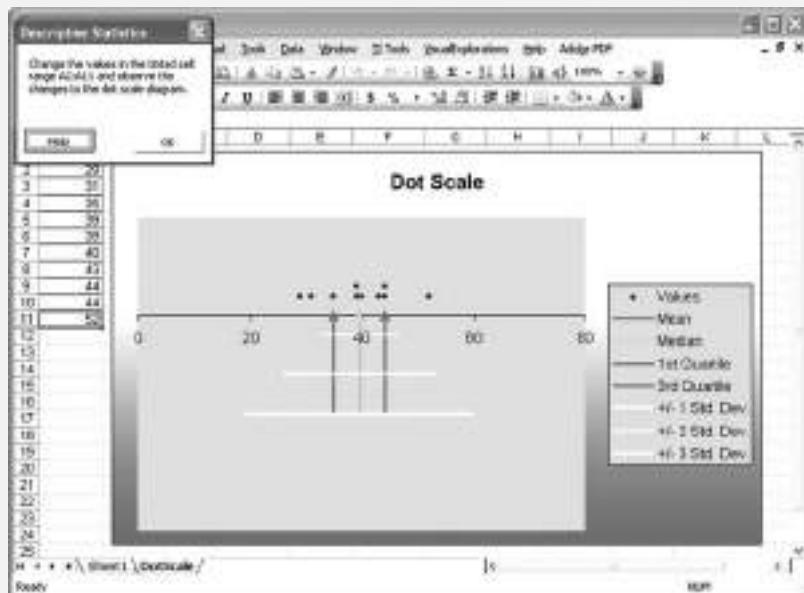
El juego de herramientas de análisis de datos de Excel genera la media, mediana, moda, desviación estandar, varianza, rango, mínimo, máximo y cuenta (tamaño de la muestra) en una sola hoja de trabajo, todos ellos analizados en esta sección. Además, Excel calcula el error estandar, lo mismo que estadísticos para la curtosis y la asimetría. El **error estandar** es igual a la desviación estandar dividida por la raíz cuadrada del tamaño de la muestra, y se estudiará en el capítulo 7. La **asimetría** mide la falta de simetría en los datos, y se basa en un estadístico que está en función de las diferencias con

## EXPLORACIONES VISUALES Exploración de la estadística descriptiva

Utilice el procedimiento Exploraciones Visuales de la Estadística Descriptiva para observar el efecto que tiene el cambio de valores en los datos sobre las medidas de tendencia central, variación y forma. Abra la macro de trabajo **Visual Explorations.xla** y seleccione **Visual Explorations → Descriptive Statistics** en la barra de herramientas de Excel. Lea las instrucciones del cuadro que aparece (vea la ilustración que se muestra a continuación) y dé clic en **OK** para examinar el diagrama de puntos correspondiente a la muestra de 10 tiempos

necesarios para arreglarse que utilizará a lo largo de este capítulo.

Experimente introduciendo un valor extremo como 10 minutos en una de las celdas de la columna A. ¿Qué medidas se ven afectadas por este cambio? ¿Cuáles no? Puede alternar entre los diagramas "previo" y "posterior" presionando repetidamente **Ctrl+Z** (deshacer) seguido de **Ctrl+Y** (rehacer) como ayuda para observar los cambios provocados por un valor extremo en el diagrama.



respecto a la media *elevadas al cubo*. Un valor de asimetría de cero indica una distribución simétrica. La *curtosis* mide la concentración relativa de valores en el centro de la distribución al compararlos con las colas y se basa en las diferencias con respecto a la media elevadas a la cuarta potencia. Esta medida no se analiza en el presente texto (vea la referencia 2).

A partir de la figura 3.2 de la página 90, los resultados de estadística descriptiva en Excel para el rendimiento de los fondos en 2003, con base en su nivel de riesgo, parecen mostrar ligeras diferencias para los tres niveles de riesgo en su rendimiento porcentual de 2003. Los fondos de alto riesgo tienen una media y una mediana ligeramente mayores que los de riesgo bajo y medio. Existe muy poca diferencia entre las desviaciones estándar de los tres grupos.

### Resultados de la estadística descriptiva en Minitab

Para la estadística descriptiva, Minitab calcula el tamaño de la muestra (etiquetado como N), media, mediana, desviación estándar (etiquetada StDev), mínimo, máximo, coeficiente de variación (etiquetado CoefVar), primer y tercer cuartiles, rango y rango intercuartil (etiquetado IQR), todos analizados en esta sección.

A partir de la figura 3.3 de la página 90, los resultados de estadística descriptiva en Minitab para el rendimiento de los fondos en 2003, con base en su riesgo, parecen registrar ligeras diferencias del rendimiento porcentual en 2003 para los tres niveles de riesgo. Los fondos de alto riesgo tienen media, mediana y cuartiles ligeramente superiores a los de riesgo bajo y medio. Existe muy poca diferencia en las desviaciones estándar o el rango intercuartil de los tres grupos.

**FIGURA 3.2**

Estadística descriptiva en Excel para el rendimiento de los fondos en 2003 con base en su nivel de riesgo.

	A	B	C	D
1	<b>Descriptive Statistics of 2003 Return by Risk</b>			
2		Low	Average	High
3	<b>Mean</b>	41.36207	42.96304	45.99412
4	<b>Standard Error</b>	1.631596	2.045318	3.117886
5	<b>Median</b>	40.25	41	44.5
6	<b>Mode</b>	35.8	37.5	#N/A
7	<b>Standard Deviation</b>	12.42586	13.87202	12.85537
8	<b>Sample Variance</b>	154.402	192.433	165.2606
9	<b>Kurtosis</b>	-0.09275	-0.33478	0.711316
10	<b>Skewness</b>	0.358427	0.586116	-0.55254
11	<b>Range</b>	55.4	58.3	51.6
12	<b>Minimum</b>	15.8	19.7	14.9
13	<b>Maximum</b>	71.2	78	66.5
14	<b>Sum</b>	2399	1976.3	781.9
15	<b>Count</b>	58	46	17

**FIGURA 3.3**

Estadística descriptiva en Minitab para el rendimiento de los fondos en 2003 con base en su nivel de riesgo.

Results for: MUTUALFUND52004.MTW									
Descriptive Statistics: Return 2003									
Variable	Risk	N	Mean	StDev	CoefVar	Minimum	Q1	Median	Q3
Return 2003	average	46	42.56	13.67	32.29	19.70	32.63	41.00	49.83
	high	17	45.59	12.66	27.05	14.90	38.25	40.50	57.05
	low	53	40.26	12.43	30.04	15.80	31.23	40.25	48.95
Variable	Risk	Maximum	Range	Q3					
Return 2003	average	78.00	58.30	17.20					
	high	66.50	51.50	18.50					
	low	71.20	56.40	17.73					

## PROBLEMAS PARA LA SECCIÓN 3.1

### Aprendizaje básico

**ASISTENCIA de PH Grade** **3.1** A continuación se encuentra un conjunto de datos procedente de una muestra de  $n = 5$ :

7 4 9 8 2

- Calcule la media, la mediana y la moda.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Calcule las puntuaciones Z. ¿Existe algún valor extremo?
- Describa la forma del conjunto de datos.

**ASISTENCIA de PH Grade** **3.2** A continuación aparece un conjunto de datos procedente de una muestra de  $n = 6$ :

7 4 9 7 3 12

- Calcule la media, la mediana y la moda.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.

- Calcule las puntuaciones Z. ¿Existe algún valor extremo?

- Describa la forma del conjunto de datos.

**ASISTENCIA de PH Grade** **3.3** A continuación aparece un conjunto de datos procedente de una muestra de  $n = 7$ :

12 7 4 9 0 7 3

- Calcule la media, la mediana y la moda.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Describa la forma del conjunto de datos.

**ASISTENCIA de PH Grade** **3.4** A continuación aparece un conjunto de datos procedente de una muestra de  $n = 5$ :

7 -5 -8 7 9

- Calcule la media, la mediana y la moda.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Describa la forma del conjunto de datos.

**ASISTENCIA de PH Grade** **3.5** Suponga que la tasa de rendimiento de una acción en particular durante los dos últimos años fue del 10 y del 30%. Calcule la media geométrica de la tasa de rendimiento (*Nota:* Una tasa de rendimiento del 10% se registra como 0.10 y una del 30% como 0.30).

## Aplicación de conceptos

**Puede resolver los problemas 3.6 a 3.20 manualmente o en Excel, Minitab o SPSS.**

**ASISTENCIA de PH Grade** **3.6** El gerente de operaciones de una fábrica de llantas quiere comparar el diámetro interno real de dos tipos de neumáticos, que se espera sean de 575 milímetros en ambos casos. Se seleccionó una muestra de cinco llantas de cada tipo y se ordenaron de menor a mayor, como se aprecia a continuación:

<b>Tipo X</b>	<b>Tipo Y</b>
568 570 575 578 584	573 574 575 577 578

- a. Calcule la media, la mediana y la desviación estándar de ambos tipos de llantas.
- b. ¿Cuál tipo de llanta es de mejor calidad? Explique por qué.
- c. ¿Qué efecto tendría en sus respuestas a los incisos a) y b) si el último valor del tipo Y fuese 588 en lugar de 578? Explique su respuesta.

**ASISTENCIA de PH Grade** **3.7** Los siguientes datos representan el total de grasas en las hamburguesas y productos de pollo de una muestra tomada de cadenas de comida rápida. **FAST-FOOD**

<i>Hamburguesas</i>									
19	31	34	35	39	39	43			
<i>Pollo</i>									
7	9	15	16	16	18	22	25	27	33

Fuente: "Quick bites", Derechos reservados © 2000 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adoptado con autorización de Consumer Reports, marzo de 2001, 46.

Para las hamburguesas y los productos de pollo realice lo siguiente por separado:

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule la varianza, la desviación estándar, el rango, el rango intercuartil y el coeficiente de variación.
- c. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. Con base en los resultados de los incisos a) a c), ¿qué conclusiones se obtienen en relación con las diferencias en la grasa total de las hamburguesas y los productos de pollo?

**3.8** La mediana del precio de una casa en diciembre de 2003 alcanza \$173,200, un incremento del 6.7% respecto a diciembre de 2002. En todo el año, las ventas alcanzaron un récord de 6.1 millones de casas (James R. Hagerty, "Housing Prices Continue to Rise", *The Wall Street Journal*, 27 de enero, 2004, D1).

- a. Describa la forma de la distribución correspondiente al precio de las casas vendidas.
- b. ¿Por qué cree usted que el artículo informa sobre la mediana de los precios y no sobre la media?

**3.9** En el ciclo escolar 2002-2003, muchas universidades públicas de Estados Unidos elevaron sus cuotas y tarifas de manutención, como resultado de la reducción de los subsidios estatales (Mary Beth Marklein, "Public Universities Raise Tuition, Fees-and Ire", *USA Today*, 8 de agosto, 2002, 1A-2A). A continuación se representa el cambio del costo de inscripción, un dormitorio compartido y el plan de alimentación más solicitado entre los ciclos escolares 2001-2002 y 2002-2003 en una muestra de 10 universidades públicas. **COLLEGECOST**

<b>Universidad</b>	<b>Cambio en el costo (\$)</b>
University of California, Berkeley	1,589
University of Georgia, Athens	593
University of Illinois, Urbana-Champaign	1,223
Kansas State University, Manhattan	869
University of Maine, Orono	423
University of Mississippi, Oxford	1,720
University of New Hampshire, Durham	708
Ohio State University, Columbus	1,425
University of South Carolina, Columbia	922
Utah State University, Logan	308

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule la varianza, la desviación estándar, el rango, el rango intercuartil, el coeficiente de variación y las puntuaciones Z.
- c. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. Con base en los resultados de los incisos a) a c), ¿qué conclusiones se obtienen en relación con el cambio de los costos entre los ciclos escolares 2001-2002 y 2002-2003?

**3.10** Los siguientes datos **COFFEDRINK** representan las calorías y la grasa (en gramos), que contienen las raciones con 16 onzas de bebidas a base de café servidas en Dunkin' Donuts y Starbucks.

<b>Producto</b>	<b>Calorías</b>	<b>Grasa</b>
Batido de moka helado de Dunkin' Donuts (pura leche)	240	8.0
Capuchino frapé de Starbucks	260	3.5
Raspado de café "Coolata" (crema) de Dunkin' Donuts	350	22.0
Café moka exprés helado de Starbucks (pura leche y crema batida)	350	20.0
Café moka batido helado de Starbucks (con crema batida)	420	16.0
Capuchino helado de Brownie de chocolate, de Starbucks (con crema batida)	510	22.0
Crema de chocolate batido helado de Starbucks (con crema batida)	530	19.0

Fuente: "Coffee as Candy at Dunkin' Donuts and Starbucks", Derechos Reservados © 2004 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057, organización sin fines de lucro. Adaptado con autorización de Consumer Reports, junio de 2004, 9, sólo con propósitos educativos. No se autoriza su reproducción o uso comercial. [www.ConsumerReports.org](http://www.ConsumerReports.org)

Para cada una de las variables (calorías y grasa):

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule la varianza, la desviación estándar, el rango, el rango intercuartil, el coeficiente de variación y las puntuaciones Z. ¿Existe un valor atípico? Explique su respuesta.
- ¿Los datos son asimétricos? De ser así, ¿cómo?
- A partir de los resultados de los incisos a) a c), ¿qué conclusiones se obtienen en relación con las calorías y la grasa de las bebidas heladas a base de café servidas en Dunkin' Donuts y en Starbucks?

**3.11** Los siguientes datos representan el costo diario de una habitación de hotel y la renta de un automóvil en 20 ciudades estadounidenses durante una semana en octubre de 2003. **HO-TEL-CAR**

Ciudad	Hotel	Automóviles
San Francisco	205	47
Los Ángeles	179	41
Seattle	185	49
Phoenix	210	38
Denver	128	32
Dallas	145	48
Houston	177	49
Minneapolis	117	41
Chicago	221	56
St. Louis	159	41
Nueva Orleáns	205	50
Detroit	128	32
Cleveland	165	34
Atlanta	180	46
Orlando	198	41
Miami	158	40
Pittsburg	132	39
Boston	283	67
Nueva York	269	69
Washington, D.C.	204	40

Fuente: The Wall Street Journal, 10 de octubre, 2003, W4.

Para cada una de las variables (costo de hotel y costo del auto):

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule la varianza, la desviación estándar, el rango, el rango intercuartil, el coeficiente de variación y las puntuaciones Z. ¿Existe un valor extremo? Explique su respuesta.
- ¿Los datos son asimétricos? De ser así, ¿cómo?
- Con base en los resultados de los incisos a) a c), ¿qué conclusiones se obtienen en relación con el costo diario de una habitación de hotel y la renta de un automóvil?

**3.12** A continuación se indica el costo de 14 modelos de cámara digital de 3 megapixeles en una tienda especializada. **CA-MERA**

340 450 450 280 220 340 290  
370 400 310 340 430 270 380

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule la varianza, la desviación estándar, el rango, el rango intercuartil, el coeficiente de variación y las puntuaciones Z. ¿Existe un valor atípico? Explique su respuesta.
- ¿Los datos son asimétricos? De ser así, ¿cómo?
- Con base en los resultados de los incisos a) a c), ¿qué conclusiones se obtienen en relación con el precio de las cámaras digitales de 3 megapixeles en una tienda especializada durante 2003?

**3.13** Una empresa dedicada a la consultoría y desarrollo de software, ubicada en el área metropolitana de Phoenix, desarrolla programas para sistemas administrativos de cadenas de suministro, con base en la reutilización sistemática de software. En lugar de comenzar desde cero al elaborar y desarrollar nuevos sistemas de software personalizados, utiliza una base de datos que contiene componentes reutilizables que suman más de 2,000,000 de líneas de código, recopilados a lo largo de 10 años de labores continuas. Se pide a 8 analistas de la empresa que calculen la tasa de reutilización cuando se desarrolla un nuevo sistema de software. Los siguientes datos corresponden al porcentaje total de código que procede de la base de datos de reutilización y forma parte del sistema de software. **REUSE**

50.0 62.5 37.5 75.0 45.0 47.5 15.0 25.0

Fuente: M. A. Rothenberger y K. J. Dooley, "A Performance Measure for Software Reuse Projects", Decision Sciences, 30 (otoño de 1999), 1131-1153.

- Calcule la media, la mediana y la moda.
- Calcule el rango, la varianza y la desviación estándar.
- Interprete las medidas sintetizadas que se calculan en los incisos a) y b).

**3.14** Un fabricante de baterías para flashes toma una muestra de 13 baterías de la producción del día y las utiliza de manera continua hasta que se agotan. El número de horas que se utilizan hasta el momento de fallar fue: **BATTERIES**

342 426 317 545 264 451  
1,049 631 512 266 492 562 298

- Calcule la media, la mediana y la moda. Al observar la distribución de los tiempos transcurridos hasta la falla, ¿cuáles medidas de ubicación le parecen más apropiadas y cuáles menos adecuadas para utilizarlas con estos datos? ¿Por qué?
- Calcule el rango, la varianza y la desviación estándar.
- ¿Qué le recomendaría a un fabricante si quisiera anunciar que sus baterías "duran 400 horas"? (Nota: No existe una respuesta exacta para esta pregunta; se trata de decir cómo hacer precisa tal afirmación.)
- Suponga que, en lugar de 342, el primer valor fue de 1,342. Repita los incisos a) a c) utilizando este valor. Elabore un comentario sobre la diferencia de los resultados.

**3.15** Una sucursal bancaria ubicada en una zona comercial de la ciudad, desarrolló un proceso mejorado para atender a sus clientes desde la hora del almuerzo al mediodía, hasta la 1:00 PM. Se registra el tiempo de espera en minutos (definido como el tiempo transcurrido desde que el cliente se forma en la fila hasta que llega a la ventanilla del cajero) de todos los clientes

durante ese horario por una semana. Se selecciona una muestra aleatoria de 15 clientes y se tienen los siguientes resultados:

**BANK1**

4.21	5.55	3.02	5.13	4.77	2.34	3.54
3.20	4.50	6.10	0.38	5.12	6.46	6.19
3.79						

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule la varianza, la desviación estándar, el rango, el rango intercuartil, el coeficiente de variación y las puntuaciones Z. ¿Existe algún valor atípico? Explique su respuesta.
- ¿Los datos son asimétricos? De ser así, ¿cómo?
- Un cliente llega a la sucursal durante la hora del almuerzo y pregunta al gerente cuánto tendrá que esperar, éste le responde “Menos de cinco minutos, con toda seguridad”. Con base en sus resultados de los incisos a) y b), evalúe la exactitud de tal afirmación.

**3.16** Suponga que otra sucursal, ubicada en una zona residencial, también se preocupa por el tiempo de espera desde de la hora del almuerzo hasta la 1:00 PM. Se registra el tiempo de espera en minutos (definido como el tiempo transcurrido desde que el cliente se forma en la fila hasta que llega a la ventanilla del cajero) de todos los clientes durante ese horario por una semana. Se selecciona una muestra aleatoria de 15 clientes y se tienen los siguientes resultados: **BANK2**

9.66	5.90	8.02	5.79	8.73	3.82	8.01
8.35	10.49	6.68	5.64	4.08	6.17	9.91
5.47						

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule la varianza, la desviación estándar, el rango, el rango intercuartil y el coeficiente de variación. ¿Existe algún valor atípico? Explique su respuesta.
- ¿Los datos son asimétricos? De ser así, ¿cómo?
- Un cliente llega a la sucursal durante la hora del almuerzo y pregunta al gerente cuánto tendrá que esperar, éste le responde: “Menos de cinco minutos, con toda seguridad”. Con base en sus resultados de los incisos a) y b), evalúe la exactitud de tal afirmación.



**3.17** China tiene el mercado con crecimiento más rápido en ventas de automóviles de pasajeros y es el cuarto mercado más grande, detrás de Estados Unidos, Japón y Alemania. Las ventas aumentaron un 61% en 2002 y un 55% en 2003 (Peter Wonacott, “A Fear Amid China’s Car Boom”, *The Wall Street Journal*, 2 de febrero, 2004, A17). Calcule la media geométrica de la tasa de incremento. (*Sugerencia*: Denote el crecimiento del 61% como  $R_1 = 0.61$ .)

**3.18** Durante el periodo transcurrido desde 2000 hasta 2003, se observó una gran volatilidad en el valor de las acciones. Los datos que se presentan en la siguiente tabla **STOCKRETURN** representan las tasas de rendimiento total del índice industrial Dow Jones, del índice Standard & Poor’s 500, del índice Russell 2000, y del índice Wilshire 5000 de 2000 a 2003.

Año	DJIA	SP500	Russell2000	Wilshire5000
2003	25.30	26.40	45.40	29.40
2002	-15.01	-22.10	-21.58	-20.90
2001	-5.44	-11.90	-1.03	-10.97
2000	-6.20	-9.10	-3.02	-10.89

Fuente: *The Wall Street Journal*, 2 de enero, 2004.

- Calcule la tasa de rendimiento geométrica de los índices Dow Jones, Standard & Poor’s 500, Russell 2000 y Wilshire 5000.
- ¿Qué conclusiones se obtienen en relación con las tasas de rendimiento geométricas de los cuatro índices bursátiles?
- Compare los resultados del inciso b) con los de los problemas 3.19b) y 3.20b).

**3.19** Durante el periodo de 2000 a 2003, se observó una gran volatilidad en el valor de las inversiones. Los datos que se presentan en la siguiente tabla **BANKRETURN** representan la tasa de rendimiento total de un certificado de depósito a un año, de un certificado de depósito a 30 meses y de un depósito en el mercado de dinero de 2000 a 2003.

Año	A 1 año	A 30 meses	Mercado de dinero
2003	1.20	1.76	0.61
2002	1.98	2.74	1.02
2001	3.60	3.97	1.73
2000	5.46	5.64	2.09

Fuente: *The Wall Street Journal*, 2 de enero, 2004.

- Calcule la tasa de rendimiento geométrica de los certificados de depósito a un año, 30 meses y en el mercado de dinero.
- ¿Qué conclusiones se obtienen en relación con las tasas de rendimiento geométricas de los tres depósitos?
- Compare los resultados del inciso b) con los de los problemas 3.18b) y 3.20b).

**3.20** Durante el periodo de 2000 a 2003, se observó una gran volatilidad en el valor de los metales. Los datos que se presentan en la siguiente tabla **METALRETURN** representan la tasa de rendimiento total de platino, oro y plata de 2000 a 2003.

Año	Platino	Oro	Plata
2003	34.2	19.5	24.0
2002	24.5	24.5	5.5
2001	-21.3	1.2	-3.0
2000	-23.3	1.8	-5.9

Fuente: *The Wall Street Journal*, 2 de enero, 2004.

- Calcule la tasa de rendimiento geométrica de platino, oro y plata.
- ¿Qué conclusiones se obtienen en relación con las tasas de rendimiento geométricas de los tres metales?
- Compare los resultados del inciso b) con los de los problemas 3.18b) y 3.19b).

### 3.2 MEDIDAS NUMÉRICAS DESCRIPTIVAS DE UNA POBLACIÓN

En la sección 3.1 se expusieron varios *estadísticos* que describen las propiedades de la tendencia central, la variación y la forma de una *muestra*. Si su conjunto de datos representa medidas numéricas de toda una *población*, necesita calcular e interpretar los *parámetros*, medidas sintetizadas para una población. En esta sección, aprenderá sobre tres parámetros descriptivos de la población, la media poblacional, la varianza poblacional y la desviación estándar poblacional.

Como ayuda para ilustrar estos parámetros, vea primero la tabla 3.5, que contiene los cinco mayores bonos de capital (en términos de activos totales) para el primero de marzo de 2004. También se indica el rendimiento a 52 semanas de cada uno de ellos. **LARGEST BONDS**

**TABLA 3.5**

Rendimiento en 2003 de la población compuesta por los cinco mayores bonos de capital.

Fondo de capital	Rendimiento a 52 semanas (en porcentaje)
Vanguard GNMA	3.8
Vanguard Total Bond Index	6.5
Pimco Total Return Admin	7.0
Pimco Total Return Instl	7.3
America Bond Fund	12.9

Fuente: The Wall Street Journal, 25 de marzo, 2004, C2.

#### La media poblacional

La **media poblacional** se representa por medio del símbolo  $\mu$ , la letra griega *mu* minúscula. La ecuación (3.13) define a la media poblacional.

#### MEDIA POBLACIONAL

La media poblacional es la suma de los valores de la población dividida por el tamaño de la población  $N$ .

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (3.13)$$

donde

$\mu$  = media poblacional

$X_i$  =  $i$ -ésimo valor de la variable  $X$

$\sum_{i=1}^N X_i$  = sumatoria de todos los valores  $X_i$  de la población

Para calcular el rendimiento medio de la población de bonos de capital listados en la tabla 3.5, se utiliza la ecuación (3.13),

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{3.8 + 6.5 + 7.0 + 7.3 + 12.9}{5} = \frac{37.5}{5} = 7.5$$

De esta manera, el rendimiento medio en 2003 de tales bonos de capital es del 7.5%.

## Varianza y desviación estándar poblacionales

La **varianza poblacional** y la **desviación estándar poblacional** miden la variación en una población. Al igual que los estadísticos muestrales relacionados, la desviación estándar poblacional es igual a la raíz cuadrada de la varianza poblacional. El símbolo  $\sigma^2$ , que es la letra griega *sigma* minúscula elevada al cuadrado, representa la varianza poblacional y el símbolo  $\sigma$ , la misma letra griega minúscula pero sin elevar al cuadrado, representa la desviación estándar poblacional. Las ecuaciones (3.14) y (3.15) definen esos parámetros. Los denominadores de los términos de la derecha de estas ecuaciones utilizan  $N$  y no el término  $(n - 1)$  que se emplea para la varianza y la desviación estándar de las muestras [vea las ecuaciones (3.9) y (3.10) de la página 82].

### VARIANZA POBLACIONAL

La varianza poblacional es la suma de las diferencias con respecto a la media de la población elevada al cuadrado y dividida por el tamaño de la población  $N$ .

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad (3.14)$$

donde  $\mu$  = media poblacional

$X_i$  =  $i$ -ésimo valor de la variable  $X$

$\sum_{i=1}^N (X_i - \mu)^2$  = sumatoria de todas las diferencias entre los valores  $X_i$  y  $\mu$ , elevadas al cuadrado

### DESVIACIÓN ESTÁNDAR POBLACIONAL

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (3.15)$$

Para calcular la varianza poblacional correspondiente a los datos de la tabla 3.5 de la página 94, se utiliza la ecuación (3.14),

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \\ &= \frac{(3.8 - 7.5)^2 + (6.5 - 7.5)^2 + (7.0 - 7.5)^2 + (7.3 - 7.5)^2 + (12.9 - 7.5)^2}{5} \\ &= \frac{13.69 + 1.00 + 0.25 + 0.04 + 29.16}{5} \\ &= \frac{44.14}{5} = 8.828 \end{aligned}$$

De esta forma, la varianza de los rendimientos es de 8.828 unidades porcentuales de rendimiento al cuadrado. Las unidades cuadradas hacen que la varianza sea difícil de interpretar. Debe utilizarse la desviación estándar, que emplea las unidades originales de los datos (rendimiento porcentual). A partir de la ecuación (3.15),

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} = \sqrt{8.828} = 2.97$$

Por lo tanto, el rendimiento típico en 2003 difiere de la media de 7.5 en aproximadamente 2.97. Esta enorme variación sugiere que los grandes bonos de capital tienen resultados muy distintos.

### La regla empírica

En la mayoría de los conjuntos de datos, una gran parte de los valores tienden a agruparse en algún lugar cercano a la mediana. En los conjuntos de datos asimétricos a la derecha, el agrupamiento se presenta a la izquierda de la media, es decir en un valor menor que la media. En los conjuntos de datos asimétricos a la izquierda, el agrupamiento se presenta a la derecha de la media, es decir en un valor mayor que la media. En los conjuntos de datos simétricos, donde la mediana y la media son iguales, con frecuencia los valores tienden a agruparse alrededor de la media y la mediana, generando una distribución con forma de campana. En las distribuciones de esta clase, utilizar la **regla empírica** permite examinar la variabilidad:

- Aproximadamente el 68% de los valores se encuentran a una distancia de  $\pm 1$  desviación estándar de la media.
- Aproximadamente el 95% de los valores se encuentran a una distancia de  $\pm 2$  desviaciones estándar de la media.
- Aproximadamente el 99.7% se encuentran a una distancia de  $\pm 3$  desviaciones estándar de la media.

La regla empírica ayuda a medir cómo se distribuyen los valores por encima y debajo de la media. Esto permite identificar los valores atípicos cuando se analiza un conjunto de datos numéricos. La regla empírica implica que, en las distribuciones con forma de campana, aproximadamente sólo uno de cada 20 valores estará alejado de la media más allá de dos desviaciones estándar en cualquier dirección. Por regla general, los valores que no se encuentran en el intervalo  $\mu \pm 2\sigma$  se consideran como posibles atípicos. Esta regla también implica que sólo alrededor de tres de cada 1,000 estarán alejados de la media más allá de tres desviaciones estándar. Por lo tanto, casi siempre se consideran como extremos los valores que no se encuentran en el intervalo  $\mu \pm 3\sigma$ . En los conjuntos de datos con mucha asimetría, o en los que por alguna otra razón no tienen forma de campana, en lugar de la regla empírica se debe aplicar la regla de Chebyshev, que se explica en la página 97.

### EJEMPLO 3.12

#### USO DE LA REGLA EMPÍRICA

La cantidad media de llenado de una población integrada por 12 latas de gaseosa es de 12.06 onzas, con una desviación estándar de 0.02. También se sabe que esta población tiene forma de campana. Describa la distribución de la cantidad de llenado de las latas. ¿Existe una gran probabilidad de que una lata tenga menos de 12 onzas de gaseosa?

**SOLUCIÓN**  $\mu \pm \sigma = 12.06 \pm 0.02 = (12.04, 12.08)$

$$\mu \pm 2\sigma = 12.06 \pm 2(0.02) = (12.02, 12.10)$$

$$\mu \pm 3\sigma = 12.06 \pm 3(0.02) = (12.00, 12.12)$$

Utilizando la regla empírica, aproximadamente el 68% de las latas tendrán entre 12.04 y 12.08 onzas, aproximadamente el 95% tendrá entre 12.02 y 12.10 onzas, y aproximadamente el 99.7% tendrá entre 12.00 y 12.12 onzas. Así que es muy poco probable que una lata tenga menos de 12 onzas.

## La regla de Chebyshev

La **regla de Chebyshev** (referencia 1) establece que para todo conjunto de datos, independientemente de su forma, el porcentaje de valores que se encuentran a una distancia de  $k$  desviaciones estándar o menos de la media, debe ser por lo menos igual a

$$(1 - 1/k^2) \times 100\%$$

Puede usar esta regla para todo valor de  $k$  mayor que 1. Considere una  $k = 2$ . La regla de Chebyshev establece que al menos  $[1 - (1/2)^2] \times 100\% = 75\%$  de los valores deben estar dentro de  $\pm 2$  desviaciones estándar de la media.

La regla de Chebyshev es muy general y se aplica a cualquier tipo de distribución. La regla señala *por lo menos* el porcentaje de valores que quedan dentro de una distancia dada de la media. Sin embargo, si el conjunto de datos tiene una forma que se aproxima a la de campana, la regla empírica reflejará con mayor precisión la mayor concentración de datos cerca de la media. En la tabla 3.6 se comparan la regla empírica y la de Chebyshev.

**TABLA 3.6**

Variación de los datos con respecto a la media.

Intervalo	Porcentaje de valores encontrados en intervalos alrededor de la media	
	Chebyshev (para toda distribución)	Regla empírica (distribución con forma de campana)
$(\mu - \sigma, \mu + \sigma)$	Al menos 0%	Aproximadamente 68%
$(\mu - 2\sigma, \mu + 2\sigma)$	Al menos 75%	Aproximadamente 95%
$(\mu - 3\sigma, \mu + 3\sigma)$	Al menos 88.89%	Aproximadamente 99.7%

### EJEMPLO 3.13

#### USO DE LA REGLA DE CHEBYSHEV

Como en el ejemplo 3.12, la media de la cantidad de llenado de una población integrada por 12 latas de gaseosa es de 12.06 onzas y una desviación estándar de 0.02. Sin embargo, no se conoce la forma de la población y no es posible suponer que tiene forma de campana. Describa la distribución de la cantidad de llenado de las latas. ¿Existe una gran probabilidad de que una lata tenga menos de 12 onzas de gaseosa?

**SOLUCIÓN**  $\mu \pm \sigma = 12.06 \pm 0.02 = (12.04, 12.08)$

$$\mu \pm 2\sigma = 12.06 \pm 2(0.02) = (12.02, 12.10)$$

$$\mu \pm 3\sigma = 12.06 \pm 3(0.02) = (12.00, 12.12)$$

Como la distribución posiblemente sea asimétrica, no es pertinente utilizar la regla empírica. Usando la regla de Chebyshev no se puede decir algo sobre el porcentaje de latas que tienen entre 12.04 y 12.08 onzas. Es posible determinar que al menos el 75% de las latas tendrán entre 12.02 y 12.10 onzas, y que por lo menos el 88.89% tendrán entre 12.00 y 12.12 onzas. Por lo tanto, entre 0 y 11.11% de las latas tienen menos de 12 onzas.

Cuando se tienen datos muestrales, estas dos reglas permiten entender cómo se distribuyen los datos alrededor de la media. En todo caso, use el valor de  $X$  que calculó, en lugar de  $\mu$  y el que calculó para  $S$  en lugar de  $\sigma$ . Los resultados calculados empleando los estadísticos muestrales son aproximaciones, ya que utilizó estadísticos muestrales ( $\bar{X}, S$ ) y no parámetros poblacionales ( $\mu, \sigma$ ).

## PROBLEMAS PARA LA SECCIÓN 3.2

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 3.21** A continuación se presenta un conjunto de datos para una población con  $N = 10$ :

7 5 11 8 3 6 2 1 9 8

- Calcule la media poblacional.
- Calcule la desviación estándar poblacional.

ASISTENCIA  
de PH Grade

- 3.22** A continuación se presenta un conjunto de datos para una población con  $N = 10$ :

7 5 6 6 6 4 8 6 9 3

- Calcule la media poblacional.
- Calcule la desviación estándar poblacional.

### Aplicación de conceptos

AUTO  
Examen

- 3.23** Los siguientes datos representan las declaraciones trimestrales de impuestos por ventas (en miles de dólares), correspondientes al periodo que finalizó en marzo de 2004, enviados al contralor del poblado Fair Lake por los 50 negocios establecidos en dicha localidad: TAX

10.3	11.1	9.6	9.0	14.5
13.0	6.7	11.0	8.4	10.3
13.0	11.2	7.3	5.3	12.5
8.0	11.8	8.7	10.6	9.5
11.1	10.2	11.1	9.9	9.8
11.6	15.1	12.5	6.5	7.5
10.0	12.9	9.2	10.0	12.8
12.5	9.3	10.4	12.7	10.5
9.3	11.5	10.7	11.6	7.8
10.5	7.6	10.1	8.9	8.6

- Calcule la media, la varianza y la desviación estándar de esta población.
- ¿Qué proporción de estos negocios tienen declaraciones trimestrales de impuestos sobre ventas dentro de  $\pm 1$ ,  $\pm 2$  o  $\pm 3$  desviaciones estándar de la media?
- Compare y encuentre las diferencias entre sus hallazgos con lo que cabría esperar de acuerdo con la regla empírica. ¿Le sorprenden los resultados obtenidos en b)?

ASISTENCIA  
de PH Grade

- 3.24** Considere una población de 1,024 fondos de inversión que invierten principalmente en empresas grandes. Usted determinó que  $\mu$ , la media del porcentaje total anual de rendimientos obtenidos por todos los fondos es 8.20 y que  $\sigma$ , la desviación estándar, es 2.75. Suponga además que determinó que el rango del porcentaje total anual va de -2.0 a 17.1 y que los cuartiles son 5.5 ( $Q_1$ ) y 10.5 ( $Q_3$ ), respectivamente. De acuerdo con la regla empírica, ¿qué porcentaje de estos fondos se espera que estén
- dentro de  $\pm 1$  desviaciones estándar de la media?
  - dentro de  $\pm 2$  desviaciones estándar de la media?

- De acuerdo con la regla de Chebyshev, ¿qué porcentaje de estos fondos se espera que estén dentro de  $\pm 1$ ,  $\pm 2$  o  $\pm 3$  desviaciones estándar de la media?
- De acuerdo con la regla de Chebyshev, se espera que al menos el 93.75% de estos fondos tengan rendimientos totales anuales entre ¿cuáles dos cantidades?

- 3.25** En la siguiente tabla ASSETS se representan los activos de cinco grandes fondos de capital, en miles de millones de dólares.

Fondo de capital	Activos (miles de millones de dólares)
Vanguard GNMA	19.5
Vanguard Total Bond Mkt. Index	16.8
Bond Fund of America A	13.7
Franklin Calif. Tax-Free Inc. A	12.8
Vanguard Short-Term Corp.	10.9

- Calcule la media de esta población constituida por los cinco bonos de capital más grandes. Interprete este parámetro.
- Calcule la varianza y la desviación estándar de esta población. Interprete estos parámetros.
- ¿Existe mucha variabilidad en los activos de los fondos de capital?

- 3.26** Los datos del archivo ENERGY contienen el consumo de energía per cápita en kilowatts·hora de cada uno de los 50 estados y el distrito de Columbia, que constituyen a Estados Unidos, durante 1999.

- Calcule la media, la varianza y desviación estándar de la población.
- ¿Qué proporción de estos estados tienen un consumo de energía promedio per cápita dentro de  $\pm 1$  desviación estándar de la media, dentro de  $\pm 2$  desviaciones estándar de la media, y dentro de  $\pm 3$  desviaciones estándar de la media?
- Compare y encuentre las diferencias entre sus hallazgos contra lo que cabría esperar de acuerdo con la regla empírica. ¿Le sorprenden los resultados obtenidos en b)?
- Eliminando los datos correspondientes al distrito de Columbia en los incisos a) a c), ¿cómo cambian los resultados?

- 3.27** Los datos en el archivo DOWRETURN muestran el rendimiento anualizado de 10 años (1994-2003) correspondiente a 30 empresas incluidas en el Dow Jones Industrials.

- Calcule la media de esta población. Interprete este número.
- Calcule la varianza y la desviación estándar de esta población. Interprete la desviación estándar.
- Utilice la regla empírica o la de Chebyshev, la que resulte apropiada, para explicar aún más la variación de este conjunto de datos.
- Utilizando los resultados de c), ¿existen algunos datos atípicos? Explique su respuesta.

### 3.3 ANÁLISIS EXPLORATORIO DE DATOS

En la sección 3.1 se analizaron estadísticos muestrales para datos numéricos como son las medidas de tendencia central, variación y forma. Otra manera de describir datos numéricos es mediante el análisis exploratorio de datos, que incluye el resumen de cinco números y la gráfica de caja y bigote (referencias 5 y 6).

#### Resumen de cinco números

Un **resumen de cinco números** compuesto por:

$$X_{\text{menor}} \quad Q_1 \quad \text{Mediana} \quad Q_3 \quad X_{\text{mayor}}$$

permite determinar la forma de la distribución. En la tabla 3.7 se explica cómo las relaciones entre los “cinco números” le permiten reconocer la forma del conjunto de datos.

**TABLA 3.7**

Relaciones entre el resumen de cinco números y el tipo de distribución

Comparación	Tipo de distribución		
	Asimétrico a la izquierda	Simétrico	Asimétrico a la derecha
La distancia de $X_{\text{menor}}$ a la mediana contra la distancia de la mediana a $X_{\text{mayor}}$ .	La distancia de $X_{\text{menor}}$ a la mediana es mayor que la distancia de la mediana a $X_{\text{mayor}}$ .	Ambas distancias son iguales.	La distancia de $X_{\text{menor}}$ a la mediana es menor que la distancia de la mediana a $X_{\text{mayor}}$ .
La distancia de $X_{\text{menor}}$ a $Q_1$ contra la distancia de $Q_3$ a $X_{\text{mayor}}$ .	La distancia de $X_{\text{menor}}$ a $Q_1$ es mayor que la distancia de $Q_3$ a $X_{\text{mayor}}$ .	Ambas distancias son iguales.	La distancia de $X_{\text{menor}}$ a $Q_1$ es menor que la distancia de $Q_3$ a $X_{\text{mayor}}$ .
La distancia de $Q_1$ a la mediana contra la distancia de la mediana a $Q_3$ .	La distancia de $Q_1$ a la mediana es mayor que la distancia de la mediana a $Q_3$ .	Ambas distancias son iguales.	La distancia de $Q_1$ a la mediana es menor que la distancia de la mediana a $Q_3$ .

Para la muestra de 10 tiempos necesarios para arreglarse, el menor valor es 29 minutos y el mayor es 52 minutos (vea las páginas 75 y 77). Los cálculos ya realizados en la sección 3.1 indican que la mediana = 39.5, el primer cuartil = 35, y el tercer cuartil = 44. Por lo tanto, el resumen de cinco puntos es:

$$29 \quad 35 \quad 39.5 \quad 44 \quad 52$$

La distancia de  $X_{\text{menor}}$  a la mediana ( $39.5 - 29 = 10.5$ ) es ligeramente menor que la distancia de la mediana a  $X_{\text{mayor}}$  ( $52 - 39.5 = 12.5$ ). La distancia de  $X_{\text{menor}}$  a  $Q_1$  ( $35 - 29 = 6$ ) es ligeramente menor que la distancia de  $Q_3$  a  $X_{\text{mayor}}$  ( $52 - 44 = 8$ ). De esta forma, los tiempos para arreglarse son ligeramente asimétricos a la derecha.

#### EJEMPLO 3.14

#### CÁLCULO DEL RESUMEN DE LOS CINCO NÚMEROS DEL PORCENTAJE DE RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE ALTO RIESGO PARA PEQUEÑOS CAPITALES

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72), se clasifican de acuerdo con el nivel de riesgo (bajo, medio y alto) y el tamaño del capital invertido (pequeño, mediano y gran capital). Calcule el resumen de cinco puntos del rendimiento en 2003 de los nueve fondos de inversión de alto riesgo para pequeños capitales. **MUTUALFUNDS2004**

**SOLUCIÓN**

De los cálculos previos realizados a los rendimientos en 2003 de los fondos de alto riesgo para pequeños capitales (vea las páginas 76 y 78), la mediana = 53.8, el primer cuartil = 41.7, y el tercer cuartil = 60.85. Además, el menor valor del conjunto de datos es 37.3 y el mayor es 66.5. Por lo tanto, el resumen de cinco puntos es:

37.3 41.7 53.8 60.85 66.5

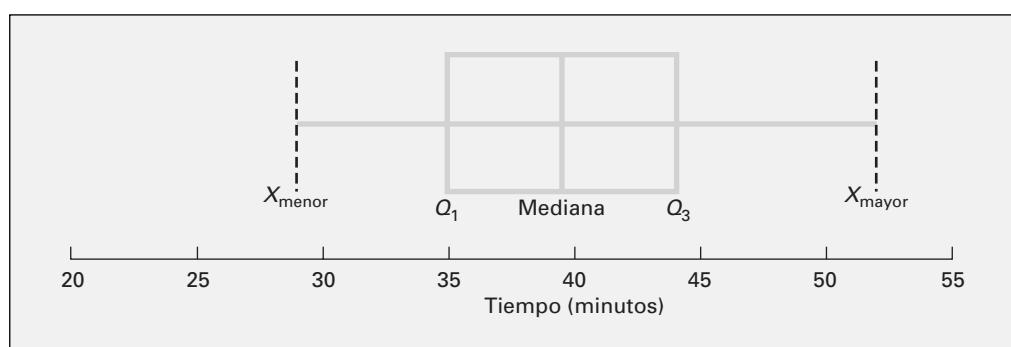
La distancia de  $X_{\text{menor}}$  a la mediana ( $53.8 - 37.3 = 16.5$ ) es mayor que la distancia de la mediana a  $X_{\text{mayor}}$  ( $66.5 - 53.8 = 12.7$ ). Esto indica asimetría a la izquierda. La distancia de  $X_{\text{menor}}$  a  $Q_1$  ( $41.7 - 37.3 = 4.4$ ) es ligeramente menor que la distancia de  $Q_3$  a  $X_{\text{mayor}}$  ( $66.5 - 60.85 = 5.65$ ). Esto indica una ligera asimetría a la derecha. Por lo tanto, los resultados son incongruentes.

**Gráfica de caja y bigote**

La **gráfica de caja y bigote** ofrece una representación visual de los datos basada en el resumen de cinco números. En la figura 3.4 se ilustra la gráfica de caja y bigote de los tiempos necesarios para arreglarse.

**FIGURA 3.4**

Gráfica de caja y bigote del tiempo necesario para arreglarse.



La línea vertical dibujada dentro de la caja representa a la mediana. La línea vertical a la izquierda de la caja representa la ubicación de  $Q_1$  y la línea vertical a la derecha de la caja representa la ubicación de  $Q_3$ . De esta forma, la caja contiene al 50% de los valores de la distribución. El 25% inferior de los datos se representa mediante una línea (es decir, un *bigote*) que une el lado izquierdo de la caja con la ubicación del menor valor,  $X_{\text{menor}}$ . De la misma manera, el 25% superior de los datos se representa mediante un bigote que une el lado derecho de la caja con la ubicación del valor mayor,  $X_{\text{mayor}}$ .

La gráfica de caja y bigote de los tiempos necesarios para arreglarse que aparece en la figura 3.4 muestra una muy ligera asimetría a la derecha, ya que la distancia entre la mediana y el valor mayor es ligeramente mayor que la distancia entre el menor valor y la mediana. El bigote derecho es un poco más largo que el izquierdo.

**EJEMPLO 3.15****GRÁFICA DE CAJA Y BIGOTE DEL RENDIMIENTO PORCENTUAL EN 2003 DE LOS FONDOS DE INVERSIÓN DE RIESGO BAJO, PROMEDIO Y ALTO**

Los 121 fondos de inversión que forman parte del escenario “Uso de la estadística” (vea la página 72) se clasifican de acuerdo con su nivel de riesgo (bajo, medio y alto) y tamaño del capital invertido (pequeño, mediano y gran capital). Construya la gráfica de caja y bigote para los rendimientos en 2003 para los fondos de inversión de riesgo bajo, promedio y alto. **MUTUALFUNDS2004**

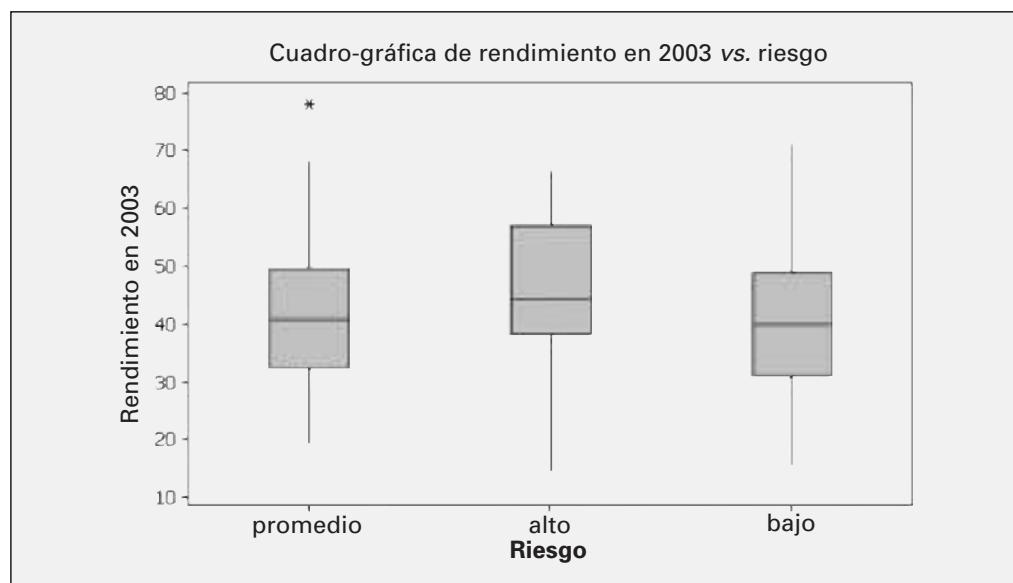
### SOLUCIÓN

En la figura 3.5 aparece la gráfica de caja y bigote de los rendimientos en 2003 para los fondos de inversión de riesgo bajo, promedio, y alto, elaborada en Minitab. Este programa muestra la gráfica de manera vertical, de inferior (bajo) a superior (alto). El asterisco (\*) de los fondos de riesgo promedio representa la presencia de valores atípicos.<sup>2</sup> La media del porcentaje de rendimiento y los cuartiles de los fondos de alto riesgo son mayores que los correspondientes a los fondos de riesgo bajo o promedio. Los fondos de riesgo promedio son asimétricos a la derecha, a causa del rendimiento extremadamente alto de uno de ellos (78). Los fondos de alto riesgo aparecen asimétricos a la izquierda por el largo bigote inferior, pero la mediana del rendimiento está más cerca del primer cuartil que del tercero. Los fondos de bajo riesgo aparecen ligeramente asimétricos a la derecha porque el bigote superior es más largo que el inferior.

<sup>2</sup>Si existen valores atípicos, los bigotes de la gráfica de caja y bigote de Minitab se extienden 1.5 veces el rango intercuartil más allá de los cuartiles o hasta el valor más alto.

**FIGURA 3.5**

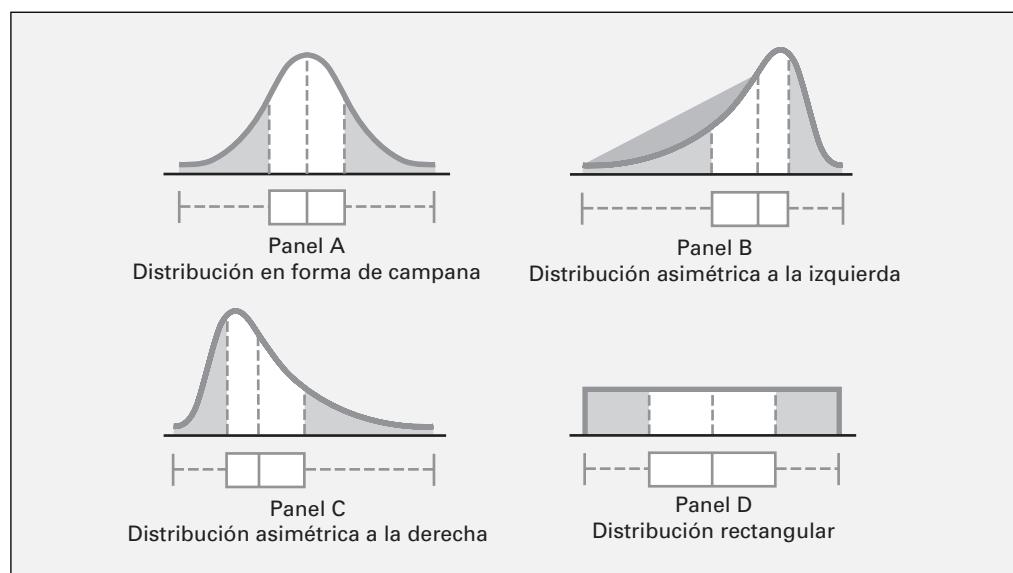
Gráfica de caja y bigote de los rendimientos en 2003, en Minitab, para los fondos de inversión de riesgo bajo, promedio y alto.



En la figura 3.6 se comprueba la relación que existe entre la gráfica de caja y bigote y el polígono de cuatro tipos distintos de distribución. (Nota: El área bajo cada polígono se divide en cuartiles que corresponden al resumen de cinco números de la gráfica de caja y bigote.)

**FIGURA 3.6**

Gráficas de caja y bigote, y sus polígonos correspondientes, de cuatro distribuciones.



Los paneles A y D de la figura 3.6 son simétricos. En estas distribuciones, la media y la mediana son iguales. Además, la longitud del bigote izquierdo es igual a la del derecho, y la línea que representa a la mediana divide la caja por la mitad.

El panel B de la figura 3.6 es asimétrico a la izquierda. Los pocos valores pequeños inclinan a la media hacia la punta izquierda. Para esta distribución asimétrica a la izquierda, la asimetría indica que existe un marcado agrupamiento de los valores en el extremo superior de la escala (es decir, el lado derecho); el 75% de todos los valores se encuentran entre el extremo izquierdo de la caja ( $Q_1$ ) y el extremo del bigote derecho ( $X_{\text{mayor}}$ ). Por lo tanto, el largo bigote izquierdo contiene al 25% más pequeño de los valores, lo que demuestra distorsión de la simetría de este conjunto de datos.

El panel C de la figura 3.6 es asimétrico a la derecha. La concentración de valores está en el extremo inferior de la escala (es decir, en el lado izquierdo de la gráfica de caja y bigote). Aquí, el 75% de todos los valores se encuentran entre el principio del bigote izquierdo ( $X_{\text{menor}}$ ) y el extremo derecho de la caja ( $Q_3$ ), y el 25% restante de los valores se encuentran dispersos a lo largo del bigote derecho, en el extremo superior de la escala.

## PROBLEMAS PARA LA SECCIÓN 3.3

### Aprendizaje básico

**ASISTENCIA** **3.28** A continuación se presenta un conjunto de datos para una muestra con  $n = 6$ :

7 4 9 7 3 12

- Elabore el resumen de cinco números.
- Construya su gráfica de caja y bigote, y describa la forma.
- Compare su respuesta del inciso b) con la del problema 3.2d) de la página 90. Analícela.

**ASISTENCIA** **3.29** A continuación se presenta un conjunto de datos para una muestra con  $n = 7$ :

12 7 4 9 0 7 3

- Elabore el resumen de cinco números.
- Realice su gráfica de caja y bigote, y describa la forma.
- Compare su respuesta del inciso b) con la del problema 3.3c) de la página 90. Analícela.

**3.30** A continuación se presenta un conjunto de datos para una muestra con  $n = 5$ :

7 -5 -8 7 9

- Elabore el resumen de cinco números.
- Construya su gráfica de caja y bigote, y describa la forma.
- Compare su respuesta del inciso b) con la del problema 3.4c) de la página 90. Analícela.

### Aplicación de conceptos

Puede resolver los problemas 3.31 a 3.36 manualmente o en Excel, Minitab o SPSS.

**ASISTENCIA**  
de PH Grade

**AUTO**  
Examen

**3.31** Un fabricante de baterías para flash fotográfico tomó una muestra de 13 baterías de la producción diaria y las utilizó

de manera continua hasta agotarlas. El número de horas que funcionaron está en el archivo **BATTERIES**

342 426 317 545 264 451  
1,049 631 512 266 492 562 298

- Elabore el resumen de cinco números.
- Construya su gráfica de caja y bigote, y describa la forma.

**3.32** Durante el ciclo escolar 2002-2003, muchas universidades estadounidenses elevaron sus cuotas y tarifas de manutención, como consecuencia de la reducción de los subsidios estatales (Mary Beth Marklein, “Public Universities Raise Tuition, Fees -and Ire”, *USA Today*, 8 de agosto, 2002, 1A-2A). A continuación se representa el cambio del costo de inscripción, un dormitorio compartido y el plan de alimentación más solicitado entre los ciclos escolares 2001-2002 y 2002-2003, para una muestra de 10 universidades públicas. **COLLEGECOST**

Universidad	Cambio en el costo (\$)
University of California, Berkeley	1,589
University of Georgia, Athens	593
University of Illinois, Urbana-Champaign	1,223
Kansas State University, Manhattan	869
University of Maine, Orono	423
University of Mississippi, Oxford	1,720
University of New Hampshire, Durham	708
Ohio State University, Columbus	1,425
University of South Carolina, Columbia	922
Utah State University, Logan	308

- Elabore el resumen de cinco números.
- Construya su gráfica de caja y bigote, y describa la forma.

**3.33** Una empresa dedicada a la consultoría y al desarrollo de software, ubicada en el área metropolitana de Phoenix, desarrolla software para sistemas administrativos de cadenas de suministro y se vale de la reutilización sistemática de software. En lugar de comenzar desde cero para elaborar y desarrollar nuevos sistemas personalizados de software, utiliza una base de datos que contiene componentes reutilizables que suman más de 2,000,000 de líneas de código, recopilados a lo largo de 10 años de actividades continuas. Se pide a ocho analistas de la empresa que calculen la tasa de reutilización cuando se desarrolla un nuevo sistema de software. Los siguientes datos corresponden al porcentaje total de código que procede de la base de datos de reutilización y forma parte del sistema de software. **REUSE**

50.0 62.5 37.5 75.0 45.0 47.5 15.0 25.0

Fuente: M. A. Rothenberger y K. J. Dooley, "A Performance Measure for Software Reuse Projects", *Decision Sciences*, 30 (Otoño de 1999), 1131-1153.

- Elabore el resumen de cinco números.
- Realice su gráfica de caja y bigote, y describa la forma de los datos.

**3.34** Los siguientes datos representan la tarifa (en dólares) por cheque devuelto de una muestra de 23 bancos, para los clientes de depósito directo que conservan un saldo de \$100 y la cuota (en dólares) mensual por manejo de cuenta, si sus cuentas no conservan el saldo mínimo requerido de \$1,500, de una muestra de 26 bancos. **BANKCOST1** **BANKCOST2**

*Tarifa por cheque devuelto*

26 28 20 20 21 22 25 25 18 25 15 20 18 20 25 25 22 30 30 30 15 20 29

*Cuota mensual por manejo de cuenta*

12 8 5 5 6 6 10 10 9 7 10 7 7 5 0 10 6 9 12 0 5 10 8 5 5 9

Fuente: "The New Face of Banking", Copyright © 2000 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, junio de 2000.

- Elabore el resumen de cinco números de la tarifa por cheque devuelto y de la cuota mensual por manejo de cuenta.
- Realice la gráfica de caja y bigote de la tarifa por cheque devuelto y de la cuota mensual por manejo de cuenta.
- ¿Qué similitudes y diferencias existen en la distribución de la tarifa por cheque devuelto y de la cuota mensual por manejo de cuenta?

**3.35** Los siguientes datos representan el total de grasas en hamburguesas y artículos de pollo tomados de una muestra de cadenas de comida rápida. **FASTFOOD**

*Hamburguesas*

19 31 34 35 39 39 43

*Pollo*

7 9 15 16 16 18 22 25 27 33 39

Fuente: "Quick Bites", Copyright © 2001 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, marzo de 2001, 46.

- Elabore el resumen de cinco puntos para las hamburguesas y para los productos de pollo.
- Construya la gráfica de caja y bigote para las hamburguesas y los productos de pollo, y describa la forma de la distribución de cada una.
- ¿Qué similitudes y diferencias existen en la distribución de hamburguesas y de productos de pollo?

**3.36** Una sucursal bancaria ubicada en una zona comercial de la ciudad desarrolló un proceso mejorado para atender a sus clientes durante la hora del almuerzo a mediodía, hasta la 1:00 PM. Durante una semana se registra el tiempo de espera en minutos (definido de manera operacional como el tiempo transcurrido desde que el cliente se forma en la fila hasta que llega a la ventanilla del cajero) de todos los clientes en ese horario. Se selecciona una muestra aleatoria de 15 clientes, y los resultados son los siguientes: **BANK1**

4.21 5.55 3.02 5.13 4.77 2.34 3.54  
3.20 4.50 6.10 0.38 5.12 6.46 6.19 3.79

Otra sucursal, ubicada en una zona residencial, también está preocupada por el horario del almuerzo de mediodía hasta la 1:00 PM. Durante una semana, se registra el tiempo de espera en minutos (definido como el tiempo transcurrido desde que el cliente se forma en la fila hasta que llega a la ventanilla del cajero) de todos los clientes en ese horario. Se selecciona una muestra aleatoria de 15 clientes, y los resultados son los siguientes: **BANK2**

9.66 5.90 8.02 5.79 8.73 3.82 8.01  
8.35 10.49 6.68 5.64 4.08 6.17 9.91 5.47

- Elabore el resumen de cinco números para tiempo de espera en ambas sucursales bancarias.
- Construya la gráfica de caja y bigote, y describa la forma de la distribución de las dos sucursales.
- ¿Qué similitudes y diferencias existen en la distribución de los tiempos de espera en ambas sucursales bancarias?

## 3.4 LA COVARIANZA Y EL COEFICIENTE DE CORRELACIÓN

En la sección 2.5, usted utilizó los diagramas de dispersión para examinar de forma visual la relación que existe entre dos variables numéricas. En esta sección, se analizan la covarianza y el coeficiente de correlación, que miden la fortaleza de la relación entre dos variables numéricas.

### La covarianza

La **covarianza** mide la fortaleza de la relación lineal entre dos variables numéricas ( $X$  y  $Y$ ). La ecuación 3.16 define la **covarianza de una muestra** y el ejemplo 3.16 ilustra su uso.

## LA COVARIANZA MUESTRAL

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \quad (3.16)$$

## EJEMPLO 3.16

## CÁLCULO DE LA COVARIANZA DE UNA MUESTRA

Considere el coeficiente de gastos y los rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. Calcule la covarianza de la muestra.

## SOLUCIÓN

La tabla 3.8 presenta el coeficiente de gastos y los rendimientos de los fondos de inversión de alto riesgo para pequeños capitales, y en la figura 3.7 aparece una hoja de Excel que calcula la covarianza de esos datos. El área de cálculos de la figura 3.7 descompone la ecuación (3.16) en un conjunto de cálculos más pequeños. A partir de la celda C17, o directamente por la ecuación (3.16), se sabe que la covarianza es 1.19738.

$$\begin{aligned} \text{cov}(X, Y) &= \frac{9.579}{9 - 1} \\ &= 1.19738 \end{aligned}$$

TABLA 3.8

Coeficiente de gastos y rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales.

## Coeficiente de gastos Rendimiento en 2003

1.25	37.3	37.3
0.72	39.2	39.2
1.57	44.2	44.2
1.40	44.5	44.5
1.33	53.8	53.8
1.61	56.6	56.6
1.68	59.3	59.3
1.42	62.4	62.4
1.20	66.5	66.5

FIGURA 3.7

Hoja de Excel que calcula la covarianza entre el coeficiente de gastos y los rendimientos en 2003 de los fondos de alto riesgo para pequeños capitales.

	A	B	C
1	Expense ratio (X)	Return 2003 (Y)	$(X - X\bar{ }) (Y - Y\bar{ })$
2	1.25	37.3	1.47078
3	0.72	39.2	7.81111
4	1.57	44.2	-1.58889
5	1.40	44.5	-0.32822
6	1.33	53.8	-0.05269
7	1.61	56.6	1.30044
8	1.68	59.3	2.53711
9	1.42	62.4	0.72444
10	1.2	66.5	-2.29489
11			
12			Calculations
13	XBar	1.353333333	=AVERAGE(A2:A10)
14	YBar	51.53333333	=AVERAGE(B2:B10)
15	n-1	8	=COUNT(A2:A10) - 1
16	Sum	9.57900	=SUM(C2:C10)
17	Covariance	1.19738	=C16/C15

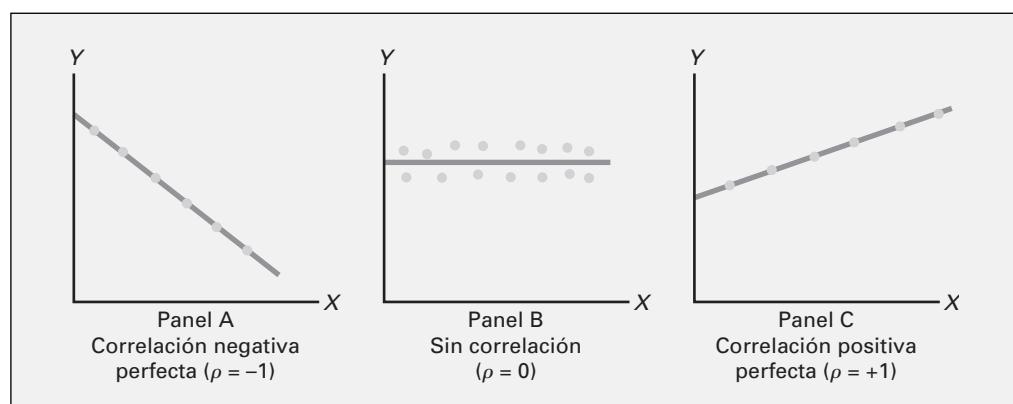
La covarianza tiene un defecto importante como medida de la relación lineal entre dos variables numéricas. Como la covarianza puede tener cualquier valor, es imposible determinar la fortaleza relativa de la relación. Para ello, es necesario calcular el coeficiente de correlación.

## Coeficiente de correlación

El **coeficiente de correlación** mide la fortaleza relativa de una relación lineal entre dos variables numéricas. Los valores del coeficiente de correlación varían desde  $-1$  para una correlación negativa perfecta, hasta  $+1$  para una correlación positiva perfecta. *Perfecta* quiere decir que si se trazaran los puntos en un diagrama de dispersión, todos ellos se podrían unir por medio de una línea recta. Al tratar con datos poblacionales para variables numéricas, se utiliza la letra griega  $\rho$  como símbolo del coeficiente de correlación. En la figura 3.8 se ilustran tres tipos diferentes de asociación entre dos variables.

**FIGURA 3.8**

Tipos de asociación entre variables.

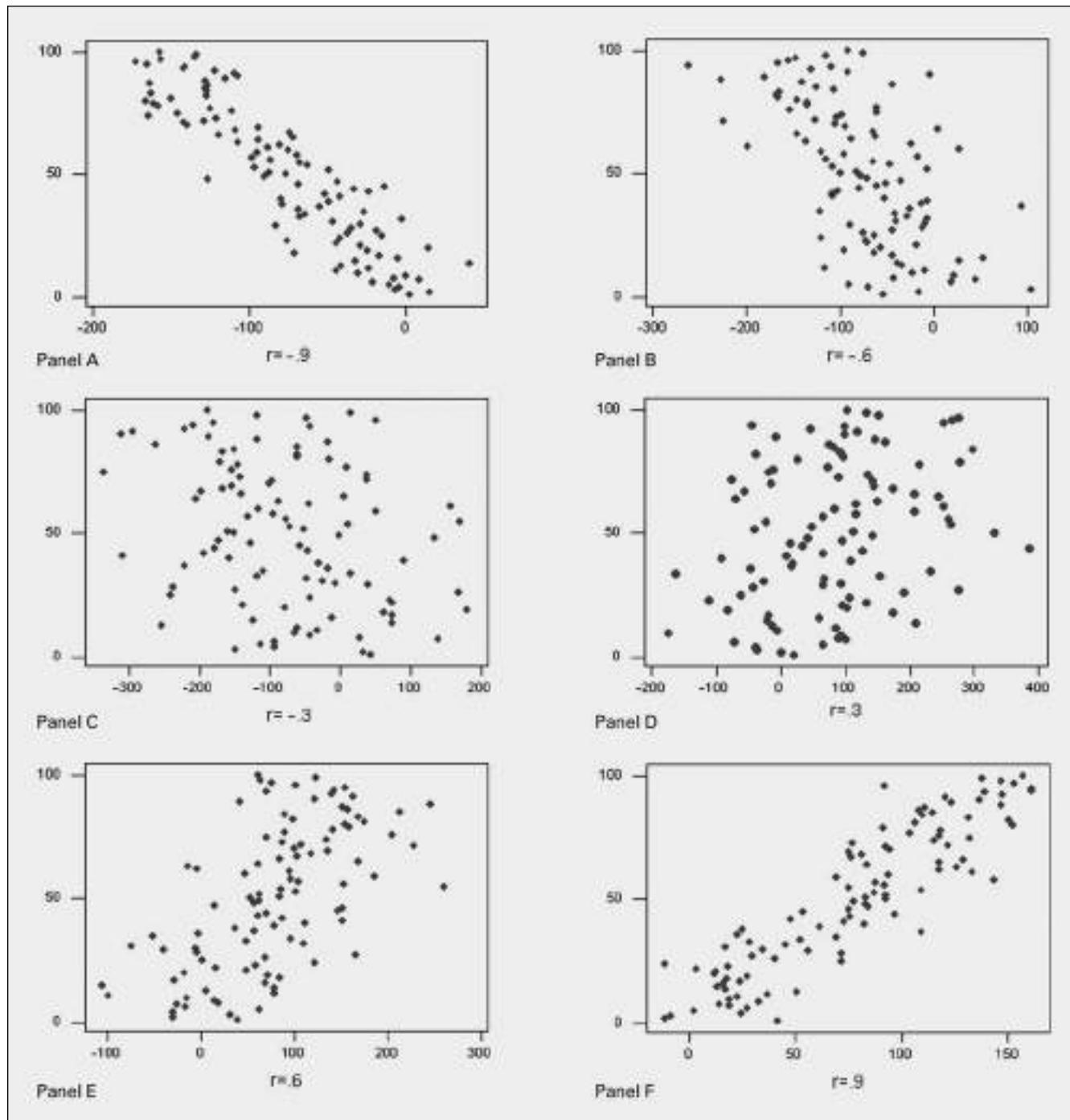


En el panel A de la figura 3.8 hay una relación lineal negativa perfecta entre  $X$  y  $Y$ . De esta manera, el coeficiente de relación  $\rho$  es igual a  $-1$ , y al aumentar  $X$ ,  $Y$  disminuye de una manera perfectamente predecible. El panel B ilustra una situación en la que no existe relación entre  $X$  y  $Y$ . En este caso, el coeficiente de correlación  $\rho$  es igual a  $0$ , y al aumentar  $X$  no existe tendencia de  $Y$  a aumentar ni disminuir. El panel C ilustra una relación positiva perfecta en la que  $\rho$  es igual a  $+1$ . En este caso,  $Y$  aumenta de una manera perfectamente predecible cuando lo hace  $X$ .

Cuando se tienen datos muestrales, se calcula el coeficiente muestral de correlación  $r$ . Al utilizar los datos de una muestra, es difícil que se tenga un coeficiente muestral de exactamente  $+1$  o  $-1$ . En la figura 3.9 de la página 106 se presentan diagramas de dispersión, con sus respectivos coeficientes muestrales de correlación  $r$  para seis conjuntos de datos, cada uno de los cuales contiene 100 valores de  $X$  y  $Y$ .

En el panel A, el coeficiente de correlación  $r$  es  $-0.9$ . Como se observa, donde los valores de  $X$  son más pequeños existe una fuerte tendencia a que los valores de  $Y$  sean grandes. De la misma forma, los valores pequeños de  $X$  tienden a hermanarse con valores pequeños en  $Y$ . No todos los datos quedan sobre una línea recta, por lo que la asociación entre  $X$  y  $Y$  no se describe como *perfecta*. Los datos del panel B tienen un coeficiente de correlación igual a  $-0.6$ , y los valores pequeños de  $X$  tienden a hermanarse con los valores grandes de  $Y$ . La relación lineal entre  $X$  y  $Y$  en el panel B no es tan fuerte como en el panel A. Así, el coeficiente de correlación en el panel B no es tan negativo como en el panel A. En el panel C, la relación lineal entre  $X$  y  $Y$  es muy débil,  $r = -0.3$ , y sólo existe una ligera tendencia de los valores pequeños de  $X$  a hermanarse con los más grandes de  $Y$ . En los paneles D a F se describen conjuntos de datos con coeficientes de correlación positivos, porque los valores pequeños de  $X$  tienden a hermanarse con los valores pequeños de  $Y$ , y los valores grandes de  $X$  tienden a asociarse con los valores grandes de  $Y$ .

En el análisis de la figura 3.9, las relaciones se describieron deliberadamente como *tendencias* y no como *causa-efecto*. Ese término se utilizó con un propósito. La sola correlación no prueba que



**FIGURA 3.9** Seis diagramas de dispersión creados con Minitab y sus respectivos coeficientes de correlación  $r$ .

existe un efecto de causalidad, es decir, que el cambio en el valor de una variable *causó* el cambio en la otra variable. Una correlación fuerte puede producirse por simple coincidencia, por el efecto de una tercera variable que no se tomó en cuenta en el cálculo, o por una relación de causa-efecto. Sería necesario realizar un análisis adicional para determinar cuál de estas tres situaciones produce verdaderamente la correlación. Por tanto, se afirma que la causalidad implica correlación, pero la correlación no implica causalidad.

La ecuación (3.17) define el **coeficiente muestral de correlación  $r$**  y el ejemplo 3.17 ilustra su uso.

## COEFICIENTE MUESTRAL DE CORRELACIÓN

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y} \quad (3.17)$$

donde  $\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$

$$S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}}$$

El ejemplo 3.17 ilustra el cálculo del coeficiente muestral de correlación  $r$  mediante la ecuación (3.17).

**EJEMPLO 3.17**

## CÁLCULO DEL COEFICIENTE MUESTRAL DE CORRELACIÓN

Considere el coeficiente de gastos y los rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales. A partir de la figura 3.10 y de la ecuación (3.17), calcule el coeficiente muestral de correlación.

**SOLUCIÓN**

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y}$$

$$= \frac{1.19738}{(0.287663)(10.554383)}$$

$$= 0.3943786$$

**FIGURA 3.10**

Hoja de Excel que calcula el coeficiente de correlación entre los gastos y los rendimientos en 2003 de los fondos de alto riesgo para pequeños capitales.

A	B	C	D	E
1	Expense ratio	Return 2003	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
2	1.25	37.3	0.0107	202.5678
3	0.72	39.2	0.4011	152.1111
4	1.57	44.2	0.0469	53.7778
5	1.4	44.5	0.0022	45.4678
6	1.33	53.8	0.0005	5.1378
7	1.61	56.6	0.0659	25.6711
8	1.68	59.3	0.1067	60.3211
9	1.42	62.4	0.0044	118.0844
10	1.2	66.5	0.0235	224.0011
11		Sum:	0.662	891.98
12				9.5790
13			Calculations	
14			XBar	1.353333333
15			YBar	51.53333333
16			$n - 1$	8
17			Covariance	1.19738
18			$S_X$	0.287662987
19			$S_Y$	10.55438298
20			$r$	0.394378596
				(formulas for range C2:E11 not shown)
				=AVERAGE(A2:A10)
				=AVERAGE(B2:B10)
				=COUNT(A2:A10) - 1
				=E11/E16
				=SQRT(C11/E16)
				=SQRT(D11/E16)
				=CORREL(A2:A10,B2:B10)
				or
				=E17/(E18 * E19)

El coeficiente de gastos y los rendimientos en 2003 de los fondos de inversión de alto riesgo para pequeños capitales están correlacionados de forma positiva. Los fondos de inversión con menores coeficientes de gastos tienden a relacionarse con los menores rendimientos en 2003. Los fondos de inversión con mayores coeficientes de gastos tienden a relacionarse con los mayores rendimientos en 2003. Esta relación es muy débil, como lo indica el coeficiente de correlación,  $r = 0.394$ .

No es posible suponer que tener un bajo coeficiente de gastos provocó los bajos rendimientos en 2003. Sólo se puede decir que eso es lo que tiende a ocurrir en la muestra. Como con todas las inversiones, los resultados del pasado no avalan los del futuro.

En resumen, el coeficiente de correlación señala la relación, o asociación, lineal entre dos variables numéricas. Cuando el coeficiente de correlación se acerca a  $+1$  o  $-1$ , es más fuerte la relación lineal entre las dos variables. Cuando el coeficiente de correlación se acerca a  $0$ , existe poca o ninguna relación lineal. El signo del coeficiente de correlación señala si los datos se correlacionan de manera positiva (es decir, los valores más grandes de  $X$  se suelen hermanar con los valores más grandes de  $Y$ ) o negativa (es decir, los valores más grandes de  $X$  se suelen hermanar con los valores más pequeños de  $Y$ ). La existencia de una correlación fuerte no implica un efecto causal. Sólo señala las tendencias presentes en los datos.

## PROBLEMAS PARA LA SECCIÓN 3.4

### Aprendizaje básico

**3.37** A continuación se presenta un conjunto de datos para una muestra con  $n = 11$  elementos:

$X$     7    5    8    3    6    10    12    4    9    15    18

$Y$     21    15    24    9    18    30    36    12    27    45    54

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- ¿Qué tan fuerte es la relación entre  $X$  y  $Y$ ? Explique su respuesta.

### Aplicación de conceptos

Puede resolver los problemas 3.38 a 3.43 manualmente o en Excel, Minitab o SPSS.

**3.38** En un artículo publicado recientemente (J. Clements, “Why Investors Should Put up to 30% of Their Stock Portfolio in Foreign Funds”, *The Wall Street Journal*, 26 de noviembre, 2003, D1) que analiza las inversiones en acciones extranjeras asegura que: el coeficiente de correlación entre el rendimiento de inversiones en acciones estadounidenses y acciones internacionales de gran capital fue de 0.80; entre acciones estadounidenses y acciones internacionales de pequeño capital fue de 0.53; entre acciones estadounidenses y bonos internacionales fue de 0.03; entre acciones estadounidenses y acciones de mercados emergentes fue de 0.71; y entre acciones estadounidenses y deuda de mercados emergentes fue de 0.58.

- ¿Qué conclusiones se obtienen sobre la fortaleza de la relación entre el rendimiento de inversiones en acciones estadounidenses y los otros cinco tipos de inversiones?
- Compare los resultados de a) con los del problema 3.39a).

**3.39** Un artículo publicado recientemente (J. Clements, “Why Investors Should Put up to 30% of Their Stock Portfolio in Foreign Funds”, *The Wall Street Journal*, 26 de noviembre, 2003, D1) que analiza las inversiones en bonos extranjeros asegura

que: el coeficiente de relación entre el rendimiento de la inversión en bonos estadounidenses y acciones internacionales de gran capital fue de  $-0.13$ ; entre bonos estadounidenses y acciones internacionales de pequeño capital fue de  $-0.18$ ; entre bonos estadounidenses y bonos internacionales fue de  $0.48$ ; entre bonos estadounidenses y acciones de mercados emergentes fue de  $-0.20$ ; y entre bonos estadounidenses y deuda de mercados emergentes fue de  $0.10$ .

- ¿Qué conclusiones se obtienen sobre la fortaleza de la relación entre el rendimiento de las inversiones en bonos estadounidenses y los otros cinco tipos de inversiones?
- Compare los resultados de a) con los del problema 3.39a).

**3.40** Los siguientes datos COFFEEDRINK representan las calorías y la grasa (en gramos) que contienen las raciones con 16 onzas de bebidas a base de café servidas en Dunkin’ Donuts y en Starbucks.

Producto	Calorías	Grasa
Batido de moka helado de Dunkin’ Donuts (pura leche)	240	8.0
Capuchino frapé de Starbucks	260	3.5
Raspado de café “Coolata” (crema) de Dunkin’ Donuts	350	22.0
Café moka exprés helado de Starbucks (pura leche y con crema batida)	350	20.0
Café moka batido helado de Starbucks (con crema batida)	420	16.0
Capuchino helado de Brownie de chocolate, de Starbucks (con crema batida)	510	22.0
77Crema de chocolate helado de Starbucks (con crema batida)	530	19.0

Fuente: “Coffee as Candy at Dunkin’ Donuts and Starbucks”, Derechos Reservados © 2004 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057, organización sin fines de lucro. Adaptado de Consumer Reports, junio de 2004, 9, sólo con propósitos educativos. No se autoriza su reproducción o uso comercial. [www.ConsumerReports.org](http://www.ConsumerReports.org)

- Calcule la covarianza de la muestra.
- Calcule el coeficiente de correlación.
- ¿Qué le parece más útil para expresar la relación que existe entre calorías y grasa: la covarianza o el coeficiente de correlación? Explique por qué.
- ¿Qué conclusiones deduce acerca de la relación entre calorías y grasa?

**3.41** Los siguientes datos representan el valor de exportaciones e importaciones de varios países en 2001: **EXPIMP**

País	Exportaciones	Importaciones
Unión Europea	874.1	912.8
Estados Unidos	730.8	1180.2
Japón	403.5	349.1
China	266.2	243.6
Canadá	259.9	227.2
Hong Kong	191.1	202.0
México	158.5	176.2
Corea del Sur	150.4	141.1
Taiwán	122.5	107.3
Singapur	121.8	116.0

Fuente: N. King y S. Miller, "Post-Iraq Influence of U.S. Faces Test at New Trade Talks", *The Wall Street Journal*, 9 de septiembre, 2003, A1.

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- ¿Qué le parece más útil para expresar la relación que existe entre exportaciones e importaciones: la covarianza o el coeficiente de correlación? Explique por qué.
- ¿Qué conclusiones puede deducir acerca de la relación entre exportaciones e importaciones?

 **3.42** Los siguientes datos **SECURITY** representan el porcentaje de traspaso durante 1998-1999 de los dispositivos de vigilancia utilizados antes de abordar en los aeropuertos, y las infracciones de seguridad detectadas por millón de pasajeros.

Ciudad	Traspaso	Infracciones
St. Louis	416	11.9
Atlanta	375	7.3
Houston	237	10.6
Boston	207	22.9
Chicago	200	6.5
Denver	193	15.2
Dallas	156	18.2
Baltimore	155	21.7
Seattle/Tacoma	140	31.5

Ciudad	Traspaso	Infracciones
San Francisco	110	20.7
Orlando	100	9.9
Washington-Dulles	90	14.8
Los Ángeles	88	25.1
Detroit	79	13.5
San Juan	70	10.3
Miami	64	13.1
Nueva York-JFK	53	30.1
Washington-Reagan	47	31.8
Honolulu	37	14.9

Fuente: Alan B. Krueger, "A Small Dose of Common Sense Would Help Congress Break the Gridlock over Airport Security", *The New York Times*, 15 de noviembre, 2001, C2.

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- ¿Qué conclusiones obtiene sobre la relación que existe entre la tasa de traspaso de los dispositivos y las infracciones de seguridad detectadas?

**3.43** Los siguientes datos **CELLPHONE** representan el tiempo en horas de uso de teléfonos móviles en modo digital y la capacidad de la batería en miliamperios.

Tiempo de uso	Capacidad de la batería	Tiempo de uso	Capacidad de la batería
4.50	800	1.50	450
4.00	1500	2.25	900
3.00	1300	2.25	900
2.00	1550	3.25	900
2.75	900	2.25	700
1.75	875	2.25	800
1.75	750	2.50	800
2.25	1100	2.25	900
1.75	850	2.00	900

Fuente: "Service Shortcomings", Copyright 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, febrero de 2002, 25.

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- ¿Qué conclusiones se obtienen sobre la relación entre la capacidad de la batería y el tiempo de uso en modo digital?
- Usted espera que los teléfonos con batería de mayor capacidad tengan un tiempo de uso superior. ¿Lo sustentan los datos?



## 3.5 ERRORES EN LAS MEDIDAS NUMÉRICAS DESCRIPTIVAS Y CONSIDERACIONES ÉTICAS

En este capítulo estudió cómo se definen las características de un conjunto de datos numéricos mediante varios estadísticos que miden las propiedades de su tendencia central, variación y forma. El siguiente paso es el análisis e interpretación de los estadísticos calculados. Su análisis es *objetivo*; su interpretación es *subjetiva*. Usted debe evitar los errores que surjan en la objetividad de su análisis o en la subjetividad de su interpretación.

El análisis de los fondos de inversión con base en el nivel de riesgo es *objetivo* y revela varios descubrimientos imparciales. Objetividad al analizar datos significa reportar las medidas numéricas descriptivas más apropiadas para un conjunto de datos determinado. Ahora que ha leído el capítulo y se ha familiarizado con varias medidas numéricas descriptivas y sus fortalezas y debilidades, ¿cómo continuará con el análisis objetivo? Como los datos se distribuyen de una manera ligeramente asimétrica, ¿no debería reportar la mediana además de la media? ¿La desviación estándar no ofrece más información sobre la propiedad de variación que el rango? ¿Debe describir al conjunto de datos como asimétrico a la derecha?

Por otra parte, la interpretación de datos es *subjetiva*. Al interpretar los descubrimientos analíticos, las personas elaboran conclusiones distintas. Todos vemos el mundo desde perspectivas diferentes. De esta manera, puesto que la interpretación de datos es subjetiva, usted debe hacerla de manera imparcial, neutral y clara.

### Aspectos éticos



En todos los análisis de datos, los aspectos éticos son de vital importancia. Como consumidor cotidiano de información, usted debe cuestionar lo que lee en periódicos y revistas, lo que escucha en la radio y la televisión, así como lo que ve en Internet. A lo largo del tiempo, se ha manifestado mucho escepticismo sobre el propósito, el enfoque y la objetividad de los estudios que se publican. Quizá ningún comentario al respecto es más representativo que la frase atribuida al famoso estadista británico del siglo XIX, Benjamin Disraeli: “Existen tres clases de mentiras: las mentiras, las mentiras de testables y la estadística”.

Las consideraciones éticas aparecen al decidir cuáles resultados incluir en un reporte. Usted debe documentar los resultados tanto buenos como malos. Además, al hacer exposiciones orales y presentar reportes escritos, debe comunicar los resultados de manera imparcial, objetiva y neutral. El comportamiento falto de ética se presenta al seleccionar de forma deliberada una medida resumida inapropiada (por ejemplo, la media de un conjunto de datos muy asimétrico), para distorsionar los hechos con el fin de respaldar una posición en particular. También es ético dejar de reportar de manera selectiva descubrimientos pertinentes, cuando éstos no respaldan una posición en particular.

## RESUMEN

Este capítulo trató sobre las medidas descriptivas. En éste y el capítulo anterior, estudió la estadística descriptiva: cómo se presentan los datos en tablas y gráficas y luego su resumen, descripción, análisis e interpretación. Al manejar los datos relacionados con los fondos de inversión, usted tuvo la oportunidad de presentar información útil mediante el uso de diagramas circulares, histogramas y otros métodos gráficos. Exploró las características del desempeño en el pasado, como la tendencia central, variabilidad y forma, utilizando medidas descriptivas numéricas como

la media, la mediana, los cuartiles, el rango, la desviación estándar y el coeficiente de correlación. En la tabla 3.9 se presenta una lista de las medidas descriptivas numéricas incluidas en este capítulo.

En el capítulo siguiente, se estudiarán los principios básicos de la probabilidad, con el fin de eliminar la brecha entre el tema de la estadística descriptiva y el de la estadística inferencial.

**TABLA 3.9**

Resumen de las medidas numéricas descriptivas.

Tipo de análisis	Datos numéricos
Describir la tendencia central, variación y forma de una variable numérica	Media, mediana, moda, cuartiles, media geométrica, rango, rango intercuartil, desviación estándar, varianza, coeficiente de variación, puntuaciones Z, gráfica de caja y bigote ( <b>secciones 3.1-3.3</b> )
Describir la relación entre dos variables numéricas	Covarianza, coeficiente de correlación ( <b>sección 3.4</b> )

# FÓRMULAS IMPORTANTES

## Media de una muestra

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (3.1)$$

## Mediana

$$\text{Mediana} = \frac{n+1}{2} \text{ valor clasificado} \quad (3.2)$$

## Primer cuartil $Q_1$

$$Q_1 = \frac{n+1}{4} \text{ valor clasificado} \quad (3.3)$$

## Tercer cuartil $Q_3$

$$Q_3 = \frac{3(n+1)}{4} \text{ valor clasificado} \quad (3.4)$$

## Media geométrica

$$\bar{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n} \quad (3.5)$$

## Media geométrica de la tasa de rendimiento

$$\bar{R}_G = [(1+R_1) \times (1+R_2) \times \cdots \times (1+R_n)]^{1/n} - 1 \quad (3.6)$$

## Rango

$$\text{Rango} = X_{\text{mayor}} - X_{\text{menor}} \quad (3.7)$$

## Rango intercuartil

$$\text{Rango intercuartil} = Q_3 - Q_1 \quad (3.8)$$

## Varianza para una muestra

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (3.9)$$

## Desviación estándar de la muestra

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (3.10)$$

## Coeficiente de variación

$$CV = \left( \frac{S}{\bar{X}} \right) 100\% \quad (3.11)$$

## Puntuaciones Z

$$Z = \frac{X - \bar{X}}{S} \quad (3.12)$$

## Media poblacional

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (3.13)$$

## Varianza poblacional

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} \quad (3.14)$$

## Desviación estándar poblacional

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (3.15)$$

## La covarianza muestral

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (3.16)$$

## Coeficiente muestral de correlación

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y} \quad (3.17)$$

# CONCEPTOS CLAVE

Atípico 86

Asimetría 88

Asimétricos positivos 88

Coeficiente de correlación 105

Coeficiente muestral de correlación 106

Coeficiente de variación 85

Covarianza 103

Covarianza de una muestra 103

Cuartiles 77

Desviación estándar 82

Desviación estándar de una muestra 82

Desviación estándar poblacional 95

Dispersión 72

Dispersión media 81

Distribución 72

Forma 72

Gráfica de caja y bigote 100

Media 73

Media aritmética 73

Media de una muestra 73

Media geométrica 79

Media poblacional 94

Mediana 75

Medidas resistentes 81

Moda 76

Puntuaciones Z	86	Regla empírica	96	Valor extremo	86
$Q_1$ : primer cuartil	77	Resumen de cinco números	99	Variación	72
$Q_2$ : segundo cuartil	77	Sesgados a la derecha	88	Varianza	82
$Q_3$ : tercer cuartil	77	Sesgados a la izquierda	88	Varianza para una muestra	82
Rango	80	Simétrica	88	Varianza poblacional	95
Rango intercuartil	81	Suma de cuadrados	82		
Regla de Chebyshev	97	Tendencia central	72		

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**3.44** ¿Cuáles son las propiedades de un conjunto de datos numéricos?

**3.45** ¿Qué expresa la propiedad tendencia central?

**3.46** ¿Cuáles son las diferencias entre media, mediana y moda, y cuáles son las ventajas y desventajas de cada una de ellas?

**3.47** ¿Cómo interpreta el primer cuartil, la mediana y el tercer cuartil?

**3.48** ¿Qué expresa la propiedad variación?

**3.49** ¿Qué mide la puntuación Z?

**3.50** ¿Cuáles son las diferencias entre las diversas medidas de la variación como rango, rango intercuartil, varianza, desviación estándar y coeficiente de variación, y cuáles son las ventajas y desventajas de cada una?

**3.51** ¿Cómo nos ayuda la regla empírica a explicar de qué maneras se agrupan y distribuyen los valores de un conjunto de datos numéricos?

**3.52** ¿En qué difieren la regla empírica y la regla de Chebyshev?

**3.53** ¿Qué expresa la propiedad forma?

**3.54** ¿En qué difieren la covarianza y el coeficiente de correlación?

### Aplicación de conceptos

*Puede resolver los problemas 3.55 a 3.61 manualmente o en Excel, Minitab, o SPSS. Le recomendamos resolver los problemas 3.62 a 3.80 con Excel, Minitab, o SPSS.*

**3.55** Una característica de calidad que resulta de interés en el proceso de llenado de bolsitas de té es el peso que contienen. Si las bolsas quedan semivacías, se presentan dos problemas. Primero, los clientes no podrían prepararse el té tan cargado como lo desean. Segundo, la empresa podría infringir las leyes de veracidad en lo descrito en la etiqueta. En este producto, el peso impreso en la etiqueta del paquete señala que, en promedio, hay 5.5 gramos de té en cada bolsa. Si la cantidad media de té en una bolsa supera ese peso, la empresa está regalando producto.

Resulta complicado introducir la cantidad exacta de té en cada bolsa, puesto que la variación en las condiciones de temperatura y humedad dentro de la fábrica, las diferencias en la densidad del té y la rápida operación de llenado que realiza la máquina (aproximadamente 170 bolsas por minuto). La siguiente tabla muestra el peso, en gramos, de una muestra compuesta por 50 bolsas de té elaboradas en una hora por una sola máquina. TEABAGS

5.65	5.44	5.42	5.40	5.53	5.34	5.54	5.45	5.52	5.41
5.57	5.40	5.53	5.54	5.55	5.62	5.56	5.46	5.44	5.51
5.47	5.40	5.47	5.61	5.53	5.32	5.67	5.29	5.49	5.55
5.77	5.57	5.42	5.58	5.58	5.50	5.32	5.50	5.53	5.58
5.61	5.45	5.44	5.25	5.56	5.63	5.50	5.57	5.67	5.36

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Interprete las medidas de tendencia central y variación dentro del contexto de este problema. ¿Por qué debería preocuparse la compañía por la tendencia central y la variación?
- Realice una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿La empresa satisface el requisito dispuesto en la etiqueta de que, en promedio, hay 5.5 gramos de té por bolsa? Si usted estuviera a cargo de este proceso, ¿qué cambios, en caso necesario, trataría de hacer con respecto a la distribución de los pesos de las bolsas individuales?

**3.56** En el estado de Nueva York las cajas de ahorro tienen permitido vender cierta clase de seguro de vida, llamado Seguro de Vida de Caja de Ahorro (SBLI, siglas en inglés para Savings Bank Life Insurance). El proceso de aprobación se compone de cada etapa de suscripción, la cual incluye una revisión de la solicitud, una consulta a la oficina de información médica, posibles peticiones de información médica adicional y exámenes médicos, así como la etapa de consolidación durante la cual se generan las pólizas y se envían al banco para su entrega. La capacidad de entregar a los clientes de manera oportuna las pólizas aprobadas resulta vital para que este servicio sea rentable para el banco. En el transcurso de un mes, se seleccionó una muestra aleatoria de 27 pólizas aprobadas, y se registró el siguiente tiempo de procesamiento total, en días: INSURANCE

73	19	16	64	28	28	31	90	60	56	31	56	22	18
45	48	17	17	17	91	92	63	50	51	69	16	17	

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- c. Elabore una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. ¿Qué le respondería usted a un cliente que entra al banco con el fin de comprar este tipo de póliza de seguros y le pregunta cuánto dura el proceso de aprobación?

**3.57** Una de las principales medidas de la calidad del servicio que brinda cualquier organización es la velocidad con la que responde a las quejas del cliente. Una gran tienda departamental, propiedad de una familia que vende muebles y pisos, incluyendo alfombras, emprendió una importante expansión durante los últimos años. En particular el departamento de pisos se amplió de dos equipos de instalación a un supervisor de instalación, un medidor, y 15 equipos de instalación. Se seleccionó una muestra de 50 quejas relacionadas con la instalación de alfombras, recibidas durante uno de los últimos años. Los siguientes datos representan el número de días transcurridos desde que se recibió la queja hasta su solución. **FURNITURE**

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- c. Elabore una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. Con base en los resultados de los incisos a) a c), si usted tuviera que informar al presidente de la empresa cuánto tendrá que esperar un cliente para ver su queja resuelta, ¿qué le diría? Explique su respuesta.

**3.58** Una empresa de manufactura produce gabinetes de acero para equipo eléctrico. El principal componente del gabinete es una canaleta que se elabora con lámina de acero calibre 14. Se produce utilizando una troqueladora de deslizamiento progresivo de 250 toneladas, que genera dos formaciones de 90 grados en el acero plano, haciendo el canal. La distancia de un lado al otro de estas formaciones resulta de especial importancia, por la impermeabilización para aplicaciones a la intemperie. La empresa necesita que la canaleta tenga una anchura de entre 8.31 y 8.61 pulgadas. A continuación encuentran las anchuras, en pulgadas, de una muestra de  $n = 49$  canaletas. **ROUGH**

8.312	8.343	8.317	8.383	8.348	8.410	8.351	8.373	8.481	8.422
8.476	8.382	8.484	8.403	8.414	8.419	8.385	8.465	8.498	8.447
8.436	8.413	8.489	8.414	8.481	8.415	8.479	8.429	8.458	8.462
8.460	8.444	8.429	8.460	8.412	8.420	8.410	8.405	8.323	8.420
8.396	8.447	8.405	8.439	8.411	8.427	8.420	8.498	8.409	

- a. Calcule la media, la mediana, el rango y la desviación estándar de la anchura. Interprete estas medidas de tendencia central y variabilidad.
- b. Elabore el resumen de cinco números.
- c. Realice su gráfica de caja y bigote y describa la forma.
- d. ¿Qué concluye sobre el número de canaletas que satisfacen las necesidades de la empresa, al medir entre 8.31 y 8.61 pulgadas de ancho?

**3.59** La empresa del problema 3.58 también fabrica aislantes eléctricos. Si los aislantes se rompen al estar en uso, es probable que ocurra un cortocircuito. Para poner a prueba la fuerza de los aislantes, se efectúa una prueba de destrucción con la finalidad de determinar cuánta fuerza se necesita para romperlos. La fuerza se mide al observar cuántas libras se aplican al aislante antes de que se rompa. A continuación se presentan los datos de 30 aislantes en este experimento: **FORCE**

1,870	1,728	1,656	1,610	1,634	1,784	1,522	1,696	1,592	1,662
1,866	1,764	1,734	1,662	1,734	1,774	1,550	1,756	1,762	1,866
1,820	1,744	1,788	1,688	1,810	1,752	1,680	1,810	1,652	1,736

- a. Calcule la media, la mediana, el rango y la desviación estándar de la variable fuerza.
- b. Interprete las medidas de tendencia central y de variabilidad del inciso a).
- c. Construya su gráfica de caja y bigote y describa la forma.
- d. ¿Qué concluye sobre la resistencia de los aislantes, si la empresa necesita una medición de al menos 1,500 libras de fuerza?

**3.60** Los problemas de una línea telefónica que impiden hacer o recibir llamadas desconciertan tanto al cliente como a la empresa telefónica. Los siguientes datos representan muestras de 20 problemas reportados a dos oficinas distintas de una empresa telefónica, y el tiempo transcurrido para resolverlos (en minutos) desde la línea del cliente: **PHONE**

*Central telefónica I Tiempo para resolver problemas (minutos)*

1.48	1.75	0.78	2.85	0.52	1.60	4.15	3.97	1.48	3.10
1.02	0.53	0.93	1.60	0.80	1.05	6.32	3.93	5.45	0.97

*Central telefónica II Tiempo para resolver problemas (minutos)*

7.55	3.75	0.10	1.10	0.60	0.52	3.30	2.10	0.58	4.02
3.75	0.65	1.92	0.60	1.53	4.23	0.08	1.48	1.65	0.72

Para ambas centrales telefónicas:

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule el rango, rango intercuartil, varianza, desviación estándar y coeficiente de variación.
- c. Elabore una gráfica de barras de lado a lado y una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. Con base en los resultados de los incisos a) a c), ¿existen algunas diferencias entre ambas centrales? Explique su respuesta.

**3.61** En muchos procesos de manufactura se utiliza el término “trabajo-en-proceso” (con frecuencia abreviado WIP, por las siglas en inglés para “work-in-process”). En una planta que produce libros, el WIP representa el tiempo que transcurre para que se doblen, junten, cosan, peguen por un extremo y encuadernen las hojas procedentes de la prensa. Los siguientes datos representan muestras de 20 libros en dos plantas de producción y el tiempo de procesamiento (definido de forma operacional como el tiempo, en días, transcurrido desde que las hojas salen de la prensa hasta que los libros se empacan en cajas) para estos trabajos. **WIP**

Planta A										
5.62	5.29	16.25	10.92	11.46	21.62	8.45	8.58	5.41	11.42	
11.62	7.29	7.50	7.96	4.42	10.50	7.58	9.29	7.54	8.92	
Planta B										
9.54	11.46	16.62	12.62	25.75	15.41	14.29	13.13	13.71	10.04	
5.75	12.46	9.17	13.21	6.00	2.33	14.25	5.37	6.25	9.71	

Para ambas plantas:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore las gráficas de barra de lado a lado y de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- Con base en los resultados de los incisos a) a c), ¿existen algunas diferencias entre ambas plantas? Explique su respuesta.

**3.62** Los datos incluidos en el archivo **CEREALS** se componen del costo monetario por onza, calorías, fibra en gramos y azúcar en gramos, de 33 cereales para desayunar.

Fuente: Obtenido de *Copyrigh 1999 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, octubre de 1999, 33-34.*

Para cada una de las variables:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué concluye en relación con el costo por onza en centavos, calorías, fibra en gramos y azúcar en gramos, de los 33 cereales para desayunar?

**3.63** Los recortes presupuestales estatales forzaron el aumento en los costos de manutención para las universidades públicas durante el ciclo escolar 2003-2004. Los datos que se encuentran en el archivo **TUITION** incluyen la diferencia en los costos de manutención entre los ciclos 2002-2003 y 2003-2004 para los alumnos procedentes del mismo estado donde se encuentra la institución y los procedentes de otros estados.

- Calcule la media, la mediana, primero y tercer cuartiles de la diferencia en los costos de manutención entre los ciclos 2002-2003 y 2003-2004 para los alumnos procedentes del mismo estado donde se encuentra la institución y los procedentes de otros estados.

- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación de la diferencia en los costos de manutención entre los ciclos 2002-2003 y 2003-2004 para los alumnos procedentes del mismo estado donde se encuentra la institución y los procedentes de otros estados.

- Elabore la gráfica de caja y bigote de la diferencia en los costos de manutención entre los ciclos 2002-2003 y 2003-2004 para los alumnos procedentes del mismo estado donde se encuentra la institución y los procedentes de otros estados. ¿Los datos son asimétricos? De ser así, ¿cómo?

- ¿Qué conclusiones obtendría en relación con la diferencia en los costos de manutención entre los ciclos 2002-2003 y 2003-2004 para los alumnos procedentes del mismo estado donde se encuentra la institución y los procedentes de otros estados?

**3.64** Las promociones de marketing, como la entrada gratis a las personas con gorra, ¿aumentan la asistencia a los juegos de la Liga Mayor de Béisbol? Un artículo publicado en *Sport Marketing Quarterly* informa sobre la efectividad de las promociones de marketing [T. C. Boyd y T. C. Krehbiel, “Promotion Timing in Major League Baseball and the Stacking Effects of Factors that Increase Game Attractiveness”, *Sport Marketing Quarterly*, 12(2003), 173-183]. El archivo de datos **ROYALS** incluye las siguientes variables para los Reales de Kansas City durante la temporada 2002:

GAME = juegos como local en el orden en que se jugaron.

ATTENDANCE = espectadores con boleto pagado en ese juego.

PROMOTION-Y = hubo promoción; N = no hubo promoción.

- Calcule la media y la desviación estándar de los espectadores con boleto pagado para los 43 juegos en los que hubo promoción y para los 37 juegos sin promoción.
- Elabore un resumen de cinco números para los 43 juegos en los que hubo promoción y para los 37 juegos sin promoción.
- Realice una representación que contenga dos gráficas de caja y bigote; una de los 43 juegos en los que hubo promoción y otra de los 37 juegos sin promoción.
- Analice los resultados de los incisos a) a c) y comente sobre la eficacia de las promociones en los juegos de los Reales durante la temporada 2002.

**3.65** Los datos incluidos en el archivo **PETFOOD2** se componen del costo por ración, tasas por lata, proteína en gramos y grasa en gramos de 97 variedades de comida seca y enlatada para perro y para gato.

Fuente: Obtenido de *Copyright 1998 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, febrero de 1998, 18-19.*

Realice lo siguiente para los cuatro tipos de comida (comida seca para perro, comida enlatada para perro, comida seca para gato y comida enlatada para gato), y para las variables costo por servicio, proteína en gramos y grasa en gramos:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.

- c. Elabore las gráficas de barras de lado a lado y la de caja y bigote, de los cuatro tipos (comida seca para perro, comida enlatada para perro, comida seca para gato y comida enlatada para gato). ¿Son asimétricos los datos de alguno de los tipos de comida? De ser así, ¿cómo?
- d. ¿Qué conclusiones obtiene en relación con las diferencias entre los cuatro tipos (comida seca para perro, comida enlatada para perro, comida seca para gato y comida enlatada para gato)?

**3.66** Un fabricante de tejas de asfalto de Boston y Vermont ofrece a sus clientes una garantía de 20 años en la mayoría de sus productos. Para determinar si una teja dura tanto como el periodo de garantía, se realiza una prueba de vida acelerada en la planta. En la prueba, realizada en un laboratorio, la teja se expone a las tensiones que recibiría en toda su vida útil de uso normal, mediante un experimento que lleva tan sólo unos minutos. En esta prueba, se cepilla repetidamente una teja durante un breve lapso, y se pesa la cantidad de gránulos (en gramos) desprendidos por el cepillado. Se espera que las tejas con menor desprendimiento duren más en uso normal que las que experimentan gran cantidad de desprendimiento. Ante esta situación, si se espera que dure tanto como el periodo de garantía, una teja no debe tener un desprendimiento superior a 0.8 gramos. El archivo **GRANULE** contiene los datos de una muestra compuesta por 170 medidas realizadas en las tejas de la empresa en Boston y 140 medidas realizadas en las tejas de Vermont.

- a. Elabore el resumen de cinco puntos para las tejas de Boston y las tejas de Vermont.
- b. Realice las gráficas barras de lado a lado y de caja y bigote para ambos tipos de teja, y describa la forma de las distribuciones.
- c. Comente sobre la capacidad de las tejas para conseguir un desprendimiento de 0.8 gramos o menos.

**3.67** Los datos del archivo **STATES** representan los resultados de la Encuesta de la Comunidad Estadounidense (American Community Survey), con una muestra de 700,000 hogares emprendida en todos los estados durante el censo de EUA del año 2000. Realice lo siguiente para las variables tiempo promedio de traslado al trabajo en minutos, porcentaje de hogares con ocho o más habitaciones, ingreso medio y porcentaje de propietarios con hipoteca, cuyos costos de vivienda superan el 30% de sus ingresos:

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- c. Realice una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. ¿Qué conclusiones obtiene en relación con el tiempo promedio de traslado al trabajo en minutos, porcentaje de hogares con ocho o más habitaciones, ingreso medio y porcentaje de propietarios con hipoteca cuyos costos de vivienda superan el 30% de sus ingresos?

**3.68** Las finanzas del béisbol han provocado mucha controversia, pues los propietarios aseguran que pierden dinero, los jugadores afirman que los propietarios ganan dinero, y los aficionados se quejan por lo costoso que resulta asistir a los juegos o verlos por televisión de paga. Además de los datos relacionados con las estadísticas del equipo durante la temporada 2001,

el archivo **BB2001** contiene las estadísticas de todos los equipos sobre precios de las entradas, índice de costo por aficionado, ingresos por entradas en temporada regular, ingresos por televisión local, radio y cable; todos los demás ingresos de operación, compensación y beneficios del jugador; datos locales y nacionales e ingresos por operaciones de béisbol. Para cada una de estas variables:

- a. Calcule la media, la mediana, primero y tercer cuartiles.
- b. Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- c. Elabore una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- d. Calcule la correlación que existe entre el número de victorias y las compensaciones y beneficios del jugador. ¿Qué tan fuerte es la relación entre estas dos variables?
- e. ¿Qué conclusiones obtiene en relación con los ingresos por entradas en temporada regular, ingresos por televisión local, radio y cable; todos los demás ingresos de operación, compensación y beneficios del jugador; datos locales y nacionales e ingresos por operaciones de béisbol?

**3.69** Los datos incluidos en el archivo **AIRCLEANERS** representan el precio, el costo anual de energía y el costo anual del filtro de unos limpiadores de aire.

- a. Calcule el coeficiente de correlación entre el precio y el costo de energía.
- b. Calcule el coeficiente de correlación entre el precio y el costo del filtro.
- c. ¿Qué conclusiones obtiene sobre la relación del costo de energía y del costo del filtro con el precio de los limpiadores de aire?

Fuente: "Portable Room Air Cleaners", Copyright © 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, febrero de 2002, 47.

**3.70** Los datos incluidos en el archivo **PRINTERS** representan el precio, la velocidad de texto, el costo de texto, el tiempo de fotografía a color y el costo de la fotografía color de unas impresoras de computadora.

- a. Calcule el coeficiente de correlación entre el precio y cada una de las siguientes características: velocidad de texto, costo de texto, tiempo de fotografía a color y costo de fotografía a color.
- b. Con base en los resultados del inciso a), ¿cree usted que alguna de las demás variables podría ser útil para pronosticar el precio de la impresora? Explique su respuesta.

Fuente: "Printers", Copyright © 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, marzo de 2002, 51.

**3.71** Usted quiere estudiar las características de los automóviles modelo 2002, en términos de las siguientes variables: millas por galón, longitud, anchura, necesidades de circunferencia de viraje, peso y capacidad del compartimiento de equipaje. **AUTO2002**

Fuente: "The 2002 Cars", Copyright © 2002 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, abril de 2002.

Para cada una de esas variables:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore una gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene en relación con los automóviles 2002?

**3.72** Consulte los datos del problema 3.71. Usted quiere comparar los vehículos utilitarios (o SUV, siglas en inglés para sports utility vehicles) con los que no son de ese tipo, en términos de millas por galón, longitud, anchura, necesidades de circunferencia de viraje, peso y capacidad del compartimiento de carga. Para cada una de esas variables, y considerando dos tipos de vehículos:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore las gráficas de barras de lado a lado y de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene en relación con las diferencias entre los SUV y los vehículos de otra clase?

**3.73** Zagat's publica las calificaciones de restaurantes en varias ciudades de Estados Unidos. El archivo **RESTRATE** contiene los datos de la calificación para la comida, decorado, servicio y precio por persona de una muestra compuesta por 50 restaurantes localizados en la ciudad de Nueva York, y 50 localizados en Long Island.

Fuente: *Zagat Survey 2002 New York City Restaurants and Zagat Survey 2002 Long Island Restaurants*.

Para los restaurantes de Nueva York y Long Island, las variables calificación de la comida, calificación del decorado, calificación del servicio y calificación del precio por persona:

- Calcule la media, la mediana, primero y tercer cuartiles.

**b.** Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.

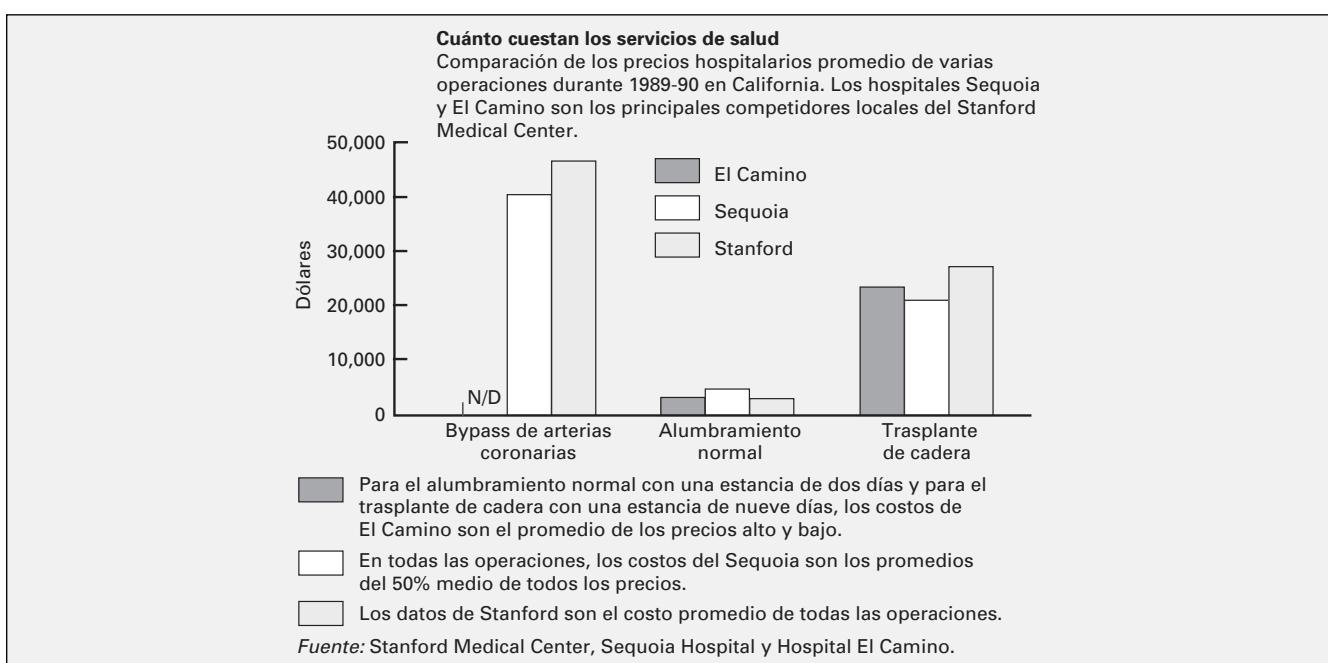
**c.** Elabore una gráfica de caja y bigote y una de barras de lado a lado de Nueva York y Long Island. ¿Son asimétricos los datos de alguna de las variables? De ser así, ¿cómo?

**d.** ¿Qué conclusiones obtiene en relación con las diferencias que existen entre los restaurantes de Nueva York y Long Island?

**3.74** Como un ejemplo del mal uso de la estadística, un artículo de Glenn Kramon ("Coaxing the Stanford Elephant to Dance", *The New York Times* Sunday Business Section, 11 de noviembre, 1990) describe que los costos del Stanford Medical Center se habían elevado más que los de la competencia ya que era más probable que brindara atención a personas indigentes, más enfermas, beneficiarios de Medicare y Medicaid, y pacientes con problemas más complejos. Se utilizó la gráfica que aparece más adelante para comparar los precios promedio en 1989 y 1990 de tres procedimientos médicos (bypass de arterias coronarias, alumbramiento normal y trasplante de cadera) en tres instituciones competidoras (El Camino, Sequoia y Stanford).

Suponga que trabaja en un centro de salud. La directora general sabe que usted está tomando un curso de estadística y le llama para analizar esto. Le dice que anoche se presentó ese artículo en el marco de una discusión de grupo, como parte de una reunión de directores generales de los centros de salud de la zona, y que uno de ellos mencionó que la gráfica era totalmente irrelevante y le pidió su opinión. Ahora ella le pide que prepare la respuesta. Usted sonríe, respira profundo y responde...

**3.75** Usted planea estudiar para su examen de estadística con un grupo de compañeros, uno de los cuales está especialmente interesado en impresionarlo. Este individuo se



ofreció a trabajar voluntariamente con Excel, Minitab o SPSS para obtener información resumida, tablas y gráficas necesarias para el conjunto de datos que contiene diversas variables numéricas y categóricas estipulado por el maestro como objeto de estudio. Se le acerca con los resultados impresos y exclama: "Lo tengo todo: —las medias, las medianas, las desviaciones estándar, las gráficas de caja y bigote, y los diagramas de pastel— de todas nuestras variables. El problema es que algunos de los resultados parecen extraños, como las gráficas de caja y bigote para género y mayores de edad, y los diagramas de pastel del índice de nivel de estudios y de la estatura. Tampoco entiendo por qué el profesor Krehbiel dice que no podemos obtener la estadística descriptiva de algunas de las variables; ¡las tengo para todo! Mira, la media de la estatura es 68.23, la media del índice de nivel de estudios es 2.76, la media del género es 1.50, la media para los mayores de edad es 4.33". ¿Cuál sería su respuesta?

### Ejercicios de reporte por escrito

**3.76** Los datos que aparecen en el archivo **BEER** representan el precio de un paquete de cerveza con 6 botellas de 12 onzas cada una, las calorías en 12 onzas líquidas, el porcentaje de contenido alcohólico en 12 onzas líquidas, el tipo de cerveza (artesanales de baja fermentación, artesanales de alta fermentación, importadas de baja fermentación, regulares y frías, y cervezas *light* y sin alcohol), y el país de origen (estadounidenses y del resto del mundo) de cada una de las 69 cervezas incluidas en la muestra.

Su tarea consiste en escribir un reporte con base en una evaluación descriptiva completa de las variables numéricas (precio, calorías y contenido alcohólico) independientemente del tipo u origen del producto. Luego realice una evaluación similar, comparando cada una de esas variables numéricas con base en el tipo de producto (artesanales de baja fermentación, artesanales de alta fermentación, importadas de baja fermentación, regulares y frías, y cervezas *light* y sin alcohol). Efectúe también una evaluación similar, para comparar y establecer las diferencias de cada una de esas variables numéricas, con base en el origen de las cervezas: las preparadas en Estados Unidos contra las del resto del mundo. Junto con su reporte debe anexar todas las tablas, los diagramas y las medidas numéricas descriptivas apropiadas.

Fuente: "Beers", Copyright © 1996 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con autorización de Consumer Reports, junio de 1996.



### PROYECTO EN EQUIPO

El archivo **MUTALFUND2004** contiene información relacionada con 12 variables a partir de una muestra de 121 fondos de inversión. Las variables son:

Fund —Nombre del fondo de inversión.

Category —Tipo de acciones que abarca el fondo de inversión: pequeño, mediano o gran capital.

Objective —Objetivo de las acciones que abarca el fondo de inversión: crecimiento o valor.

Assets —Activos en millones de dólares.

Fees —Cargos por venta (no o sí).

Expense ratio —Relación entre gastos y activos netos, en porcentaje.

2003 Return —Rendimiento en los 12 meses de 2003.

Three-year return —Rendimiento anualizado 2001 a 2003.

Five-year return —Rendimiento anualizado 1999 a 2003.

Risk —Factor de riesgo de pérdida del fondo de inversión, clasificado como bajo, medio o alto.

Best quarter —Mejor resultado trimestral 1999 a 2003.

Worst quarter —Peor resultado trimestral 1999 a 2003.

**3.77** Para la relación de gastos en porcentaje, el rendimiento en 2003, el rendimiento trianual y el rendimiento quinquenal:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore la gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene en relación con estas variables?

**3.78** Usted quiere comparar los fondos de inversión que tienen cuotas o cargos con los que no los tienen. Realice lo siguiente con cada uno de los dos grupos, para las variables relación de gastos en porcentaje, rendimiento en 2003, rendimiento trianual y rendimiento quinquenal:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore la gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene en cuanto a las diferencias que existen entre los fondos de inversión con y sin cuotas?

**3.79** Usted quiere comparar los fondos de inversión que tienen un objetivo de crecimiento con los que tienen un objetivo de valor. Realice lo siguiente con cada uno de los dos grupos, para las variables coeficiente de gastos en porcentaje, rendimiento en 2003, rendimiento trianual y rendimiento quinquenal:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore la gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene en cuanto a las diferencias que existen entre los fondos con objetivo de crecimiento y los fondos con objetivo de valor?

**3.80** Usted quiere comparar los fondos de inversión para pequeño, mediano y gran capital. Realice lo siguiente con cada uno de los tres grupos, para las variables coeficiente de gastos en porcentaje, rendimiento en 2003, rendimiento trianual y rendimiento quinquenal:

- Calcule la media, la mediana, primero y tercer cuartiles.
- Calcule el rango, el rango intercuartil, la varianza, la desviación estándar y el coeficiente de variación.
- Elabore la gráfica de caja y bigote. ¿Los datos son asimétricos? De ser así, ¿cómo?
- ¿Qué conclusiones obtiene con respecto a las diferencias que existen entre los fondos de inversión para pequeño, mediano y gran capital?

## CASO ACTUAL

### ADMINISTRACIÓN DEL SPRINGVILLE HERALD

¿Para qué variable del caso Administración del *Springville Herald* del capítulo 2 (vea la página 62) son necesarias las medidas numéricas descriptivas? Para la variable que identifique:

1. Calcule las medidas descriptivas numéricas apropiadas y elabore un diagrama de caja y bigote.

2. Identifique otra representación gráfica que resulte útil y constrúyala. ¿Qué conclusiones obtiene del hecho de que la representación no se puede hacer a partir de la gráfica de caja y bigote?

Sintetice sus hallazgos en un reporte que incluya con el estudio de la fortaleza de la tarea.

## CASO WEB

Aplique sus conocimientos sobre el uso de las medidas numéricas descriptivas a este Caso Web que es continuación del capítulo 2.

Visite de nuevo el sitio web de servicio de inversión StockTout [www.prenhall.com/Springville/StockToutHome.htm](http://www.prenhall.com/Springville/StockToutHome.htm), reexamine su datos de respaldo y luego responda lo siguiente:

1. Reexamine los datos que exploró al resolver el Caso Web del capítulo 2. ¿Es posible calcular medidas descriptivas de todas las variables? ¿Cómo respaldarían estas estadísti-

cas resumidas las demandas de StockTout? ¿Cómo influyen esas estadísticas resumidas en su percepción del registro StockTout?

2. Evalúe los métodos utilizados por StockTout para resumir los resultados de su encuesta a los clientes [www.prenhall.com/Springville/ST\\_Survey.htm](http://www.prenhall.com/Springville/ST_Survey.htm). ¿Hay algo que usted haría de otra manera para resumir estos resultados?
3. Observe que la última pregunta de la encuesta tiene menos respuestas. ¿Qué factores pueden haber limitado el número de respuestas a esa pregunta?

## REFERENCIAS

1. Kendall, M. G. y A. Stuart, *The Advanced Theory of Statistics*, vol. 1 (Londres: Charles W. Griffin, 1958).
2. *Microsoft Excel 2003* (Redmond, WA: Microsoft Corporation, 2002).
3. *Minitab Version 14* (State College, PA: Minitab Inc., 2004).
4. *SPSS Base 12.0 Brief Guide* (Upper Saddle River, NJ: Prentice Hall, 2003).
5. Tukey, J., *Exploratory Data Analysis* (Reading, MA: Addison-Wesley, 1977).
6. Velleman, P. F. y D. C. Hoaglin, *Applications, Basics, and Computing of Exploratory Data Analysis* (Boston, MA: Duxbury Press, 1981).

## Apéndice 3 Uso de software para la estadística descriptiva

### A3.1 EXCEL

#### Para la estadística descriptiva

Use el Data Analysis ToolPak. Abra la hoja de trabajo que contiene los datos que desea resumir. Seleccione **Herramientas** → **Análisis de datos**. En la lista que aparece en la ventana de diálogo Análisis de datos, seleccione **Estadística descriptiva** y dé

clic en **Aceptar**. En el cuadro de diálogo Estadística descriptiva (vea la figura A3.1), introduzca el rango de celdas de los datos en el cuadro **Rango de entrada**. Seleccione la opción **Columnas** y, si está utilizando datos ordenados como los de los archivos de Excel incluidos en el disco compacto que acompaña este libro, **Rótulos en la primera fila**. Seleccione **En una hoja nueva**, **Resumen de estadísticas**, **K-ésimo mayor** y **K-ésimo**

menor, y dé clic en **Aceptar**. Los resultados aparecen en otra hoja de trabajo.

Usted puede usar cualquiera de esas funciones de estadísticas muestrales de la hoja de trabajo con sus propias fórmulas, incluyendo PROMEDIO (para la media), MEDIANA, MODA, CUARTIL, DESVEST, VAR, MIN, MAX, SUMA, CONTAR, MAYOR o MENOR.



**FIGURA A3.1** Ventana de diálogo estadística descriptiva para el análisis de datos.

Para introducir una de esas funciones en la hoja de trabajo, seleccione una celda vacía y luego **Insertar → Función**. En el cuadro de diálogo Función, seleccione **Estadística** en la lista desplegable y luego desplácese hasta encontrar y seleccionar la función que desea utilizar. Dé clic en **Aceptar**. En la ventana de diálogo Argumentos de la función, introduzca el rango de celdas de los datos a resumir, y dé clic en **Aceptar**. (Para MAYOR y MENOR, introduzca 1 como valor de K; y para CUARTIL, introduzca 1 o 3 como valor de Cuart, según se trate del primero o tercer cuartil.) En las versiones de Excel previas a Excel 2003, puede encontrar errores en los resultados al utilizar la función CUARTIL.

### Para la gráfica de caja y bigote

Consulte la sección G.5 (Gráfica de caja y bigote) si desea que PHStat2 genere una gráfica de caja y bigote como diagrama de Excel. (No existen comandos de Excel que generen de manera directa gráficas de caja y bigote.)

### Para la covarianza

Abra el archivo de Excel **Covariance.xls**, que se muestra en la figura 3.7 de la página 104. Si desea utilizar esta hoja con otros pares de variables, siga las instrucciones en pantalla para modificar el área de la tabla. Observe en la figura 3.7 que la celda C15 contiene una fórmula que usa la función CONT. Ésta permite que Excel actualice de forma automática el valor de  $n$  cuando se modifica el tamaño del área de la tabla, y garantiza que el término  $n - 1$  siempre sea el correcto.

### Para el coeficiente de correlación

Abra el archivo de Excel **Correlation.xls**, que se ilustra en la figura 3.10 de la página 107. Si desea utilizar esta hoja con otros pares de variables, siga las instrucciones en pantalla para modificar el área de la tabla. Observe en la figura 3.10 que la celda E16 contiene una fórmula que usa la función CONT. Ésta permite que Excel actualice de forma automática el valor de  $n$  cuando se modifica el tamaño del área de la tabla, y garantiza que el término  $n - 1$  siempre sea el correcto.

Esta hoja utiliza la función CORREL para calcular el coeficiente de correlación. Como se muestra en la figura 3.10, la fórmula  $=E17/(E18 * E19)$  también se puede emplear en esta hoja para calcular el estadístico, entonces la covarianza  $S_X$  y  $S_Y$  ya aparece en la hoja.

## A3.2 MINITAB

### Cálculo de estadística descriptiva

Para generar la estadística descriptiva de los rendimientos en 2003 correspondientes a los distintos niveles de riesgo que aparecen en la figura 3.3 de la página 90, abra la hoja de trabajo **MUTUALFUNDS2004.MTW**. Seleccione **Stat → Basic Statistics → Display Descriptive Statistics**.

**Paso 1:** En la ventana de diálogo Display Descriptive Statistics (vea la figura A3.2), introduzca **C7 o Return 2003** en el cuadro de edición **Variables**. Escriba **C10 o Risk** en la ventana de editar **By variables (optional)**:



**FIGURA A3.2** Ventana de diálogo de pantalla de estadística descriptiva de Minitab.

**Paso 2:** Seleccione el botón **Statistics**. En la ventana de diálogo Display Descriptive Statistics-Statistics (vea la figura A3.3), seleccione las casillas **Mean, Standard deviation, Coefficient of variation, First quartile, Median, Third quartile, Interquartile range, Minimum, Maximum, Range y N total** (tamaño de la muestra). Dé clic en el botón **OK** para volver a la ven-

tana de diálogo Display Descriptive Statistics. Dé clic de nuevo en el botón **Aceptar** para calcular la estadística descriptiva.

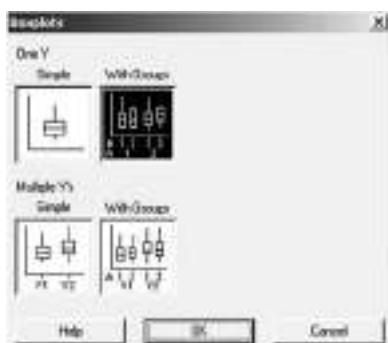


**FIGURA A3.3** Ventana de diálogo de estadística descriptiva de de Minitab.

### Uso de Minitab para elaborar una gráfica de caja y bigote

Para crear una gráfica de caja y bigote de los rendimientos en 2003 correspondientes a los distintos niveles de riesgo que aparecen en la figura 3.5 de la página 101, abra la hoja de trabajo **MUTUALFUNDS2004.MTW**. Seleccione **Graph → Boxplot**.

**Paso 1:** En la ventana de diálogo Boxplot (vea la figura A3.4) seleccione la opción **One Y With Groups** (si desea generar una gráfica de caja y bigote para un grupo, seleccione la opción **One Y Simple**). Dé clic en el botón **OK**.



**FIGURA A3.4** Ventana de diálogo Boxplots de Minitab.

**Paso 2:** En la ventana de diálogo Boxplot-One Y, With Groups (vea la figura A3.5), introduzca **C7** o **Return 2003** en el cuadro de edición **Graph variables**. Escriba **C10** o **Risk** en la ventana de editar **Categorical variables for grouping (1-4, *autolevels* test)**. Dé clic en el botón **OK**.

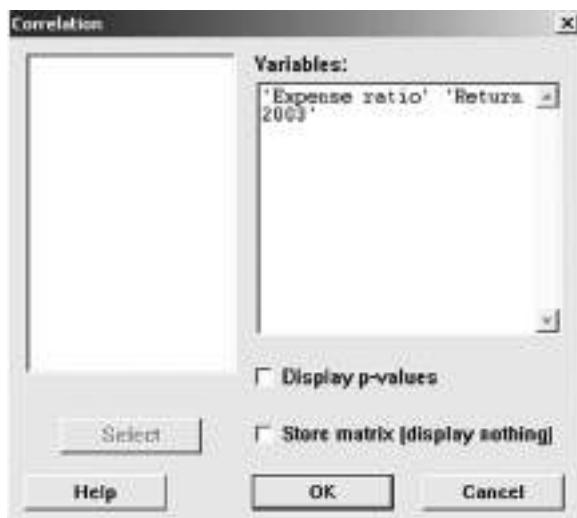


**FIGURA A3.5** Ventana de diálogo Boxplots-One Y, With Groups de Minitab.

El resultado será semejante al de la figura 3.5 de la página 101.

### Cálculo del coeficiente de correlación

Para calcular el coeficiente de correlación del coeficiente de gastos y los rendimientos en 2003 de *todos* los fondos de inversión, abra la hoja de trabajo **MUTUALFUNDS2004.MTW**. Seleccione **Stat → Basic Statistics → Correlation**. En el cuadro de diálogo Correlation (vea la figura A3.6), introduzca **C6** o **Expense ratio** y **C7** o **Return 2003**. Dé clic en el botón **OK**.



**FIGURA A3.6** Ventana de diálogo Correlation de Minitab.

# CAPÍTULO 4

## Probabilidad básica

---

**USO DE LA ESTADÍSTICA:** La empresa Consumer Electronics

### 4.1 CONCEPTOS BÁSICOS DE PROBABILIDAD

Espacios muestrales y eventos  
Tablas de contingencia y diagramas de Venn  
Probabilidad simple (marginal)  
Probabilidad conjunta  
Regla general de la adición

### 4.2 PROBABILIDAD CONDICIONAL

Cálculo de probabilidades condicionales  
Árboles de decisión  
Independencia estadística

Reglas de multiplicación  
Probabilidad marginal usando la regla general de la multiplicación

### 4.3 TEOREMA DE BAYES

### 4.4 REGLAS DE CONTEO

### 4.5 CONSIDERACIONES ÉTICAS Y PROBABILIDAD

### A.4 USO DEL SOFTWARE PARA LA PROBABILIDAD BÁSICA

*A4.1 Excel*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- Conceptos de probabilidad básica
- Probabilidad condicional
- El uso del teorema de Bayes para revisar probabilidades
- Varias reglas de conteo

# USO DE LA ESTADÍSTICA



## La empresa Consumer Electronics

Usted es gerente de marketing de la empresa Consumer Electronics. Analiza los resultados de una encuesta realizada en 1,000 hogares concerniente a sus intenciones de comprar un equipo de televisión de pantalla grande (definido como de 31 pulgadas o más) en los siguientes 12 meses. Investigaciones de este tipo reciben el nombre de estudios de intención de compra. Como seguimiento encuestará los mismos hogares 12 meses después para ver si realmente compraron el equipo. Además, a usted le interesa saber si quienes compraron la televisión de pantalla grande también compraron una televisión de alta definición (HDTV), si adquirieron un equipo reproductor de DVD en los últimos 12 meses, y si estuvieron satisfechos con la compra del equipo de televisión de pantalla grande. Algunas de las preguntas que le gustaría plantear son:

- ¿Cuál es la probabilidad de que en un hogar se planee comprar un televisor de pantalla grande el año próximo?
- ¿Cuál es la probabilidad de que en ese hogar se compre realmente un televisor de pantalla grande?
- ¿Cuál es la probabilidad de que en un hogar en el que se planea comprar un televisor de pantalla grande éste realmente se adquiera?
- Si en un hogar se planea comprar una televisión de pantalla grande, ¿cuál es la probabilidad de que la compra se realice?
- ¿El conocimiento de que en ese hogar se *planea* comprar un televisor cambia la posibilidad de predecir si ahí se *comprará* el equipo de televisión?
- ¿Cuál es la probabilidad de que el hogar donde se compra un televisor de pantalla grande se comprará un HDTV?
- ¿Cuál es la probabilidad de que en un hogar donde se compra un televisor de pantalla grande también se compre un equipo reproductor de DVD?
- ¿Cuál es la probabilidad de que un hogar donde se compra un televisor de pantalla grande estará satisfecho de su compra?

Las respuestas a estas y otras preguntas le ayudarán a desarrollar futuras estrategias de ventas y de marketing. Por ejemplo, ¿las campañas de venta para los equipos de televisión de pantalla grande deberían enfocarse en aquellos clientes que manifiestan su intención de comprar? ¿Se persuade con mayor facilidad a los individuos dispuestos a comprar un televisores de pantalla grande de comprar uno de alta definición y/o un DVD?

Los principios de la probabilidad ayudan a unir los mundos de la estadística descriptiva y de la estadística inferencial. Leer este capítulo le ayudará a aprender sobre los diferentes tipos de probabilidades y a revisarlos a la luz de nueva información. Estos temas son fundamentales para la distribución de la probabilidad, el concepto de la esperanza matemática y las distribuciones binomiales y de Poisson (temas que se estudiarán en el capítulo 5).

### 4.1 CONCEPTOS BÁSICOS DE PROBABILIDAD

¿Qué entendemos por la palabra *probabilidad*? Una **probabilidad** es un valor numérico que representa la oportunidad o posibilidad de que un evento en particular ocurra, tal como el aumento en el precio de una acción, un día lluvioso, una unidad de producción no conformada, o que caiga el cinco al lanzar un dado. En todos estos casos, la probabilidad es una proporción o fracción cuyo valor

varía entre 0 y 1 inclusive. Un evento que no tiene oportunidad de ocurrir (por ejemplo, un **evento imposible**) tiene una probabilidad de 0. Un evento que ocurrirá con toda seguridad (es decir, un **evento seguro**) tiene una probabilidad de 1. Existen tres aproximaciones sujetas a la probabilidad:

- probabilidad clásica *a priori*
- probabilidad clásica empírica
- probabilidad subjetiva

En una **probabilidad clásica *a priori***, la probabilidad de éxito se basa en el conocimiento previo del proceso implicado. En el caso más simple, en el que cada resultado es igualmente probable, la oportunidad de ocurrencia de un evento se define en la ecuación 4.1.

### PROBABILIDAD DE OCURRENCIA

$$\text{Probabilidad de ocurrencia} = \frac{X}{T} \quad (4.1)$$

donde

$X$  = número de formas en las que el evento ocurre

$T$  = número total de resultados posibles

Considere un mazo de cartas estándar con 26 cartas rojas y 26 cartas negras. La probabilidad de seleccionar una carta negra es de  $26/52 = 0.50$ , puesto que hay  $X = 26$  cartas negras y  $T = 52$  cartas en total. ¿Qué indica esta probabilidad? Si se reemplaza cada carta después de haberla seleccionado, ¿significa que una de las dos siguientes cartas será negra? No, porque usted no puede decir con certeza lo que sucederá en las selecciones posteriores. Sin embargo, puede decir que a la larga, si este proceso de selección se repite continuamente, la proporción de cartas negras seleccionadas se aproximará a 0.50.

### EJEMPLO 4.1

### ENCONTRAR PROBABILIDADES A PRIORI

Un dado estándar tiene seis caras. Cada cara contiene uno, dos, tres, cuatro, cinco o seis puntos. Si usted tira el dado, ¿cuál es la probabilidad de que caiga la cara de cinco puntos?

**SOLUCIÓN** Cada cara tiene la misma posibilidad de ocurrir. Como hay seis caras, la probabilidad de obtener la cara con cinco puntos es de  $\frac{1}{6}$ .

Los ejemplos descritos usan el punto de vista de la probabilidad clásica *a priori* porque el número de formas en las que un evento puede ocurrir y el número total de resultados posibles se conocen por la composición del mazo de cartas o de las caras del dado.

En el punto de vista de la **probabilidad clásica empírica**, los resultados se basan en datos observados, no en un conocimiento previo del proceso. Ejemplos de este tipo de probabilidad son la proporción de individuos en el escenario “Uso de la estadística” que realmente comprén la televisión, la proporción de votantes registrados que optan por un determinado candidato político, o la proporción de alumnos que tienen un empleo de medio tiempo. Por ejemplo, si usted realiza una encuesta a alumnos, y el 60% de ellos afirman que tienen un trabajo de medio tiempo, entonces hay una probabilidad de 0.60 de que un alumno en particular tenga un trabajo de medio tiempo.

El tercer punto de vista de la probabilidad, la **probabilidad subjetiva**, se distingue de los otros dos en que la probabilidad subjetiva difiere de persona a persona. Por ejemplo, tal vez el equipo de desarrollo para un nuevo producto asigne una probabilidad de 0.6 a la oportunidad de éxito para el producto, mientras que el presidente de la empresa es menos optimista y asigna una probabilidad de 0.3. La asignación de probabilidades subjetivas a diferentes resultados generalmente se basa en una combinación de las experiencias pasadas del individuo, la opinión personal y el análisis de una situación particular. La probabilidad subjetiva es particularmente útil al tomar decisiones en situaciones en las que no es posible usar la probabilidad clásica *a priori* o la probabilidad clásica empírica.

## Espacios muestrales y eventos

Los elementos básicos de la teoría de probabilidad son los resultados individuales de una variable que se somete a estudio. Para entender las probabilidades es necesario que comprenda las siguientes definiciones.

Cada posible resultado de una variable es un **evento**.

Un **evento simple** se describe por sus características singulares.

Por ejemplo, cuando lanza una moneda al aire, los dos posibles resultados son cara o cruz. Cada uno de éstos representa un evento sencillo. Cuando tira un dado estándar de seis lados, en el que las seis caras del dado contienen uno, dos, tres, cuatro, cinco o seis puntos, hay seis eventos sencillos posibles. Un evento puede ser uno de estos eventos simples, un conjunto de ellos o un subconjunto de todos ellos. Por ejemplo, el evento de un *número par de puntos* consiste en tres eventos sencillos (por ejemplo, dos, cuatro o seis puntos).

Un **evento conjunto** es un evento que tiene dos o más características.

Sacar dos caras al lanzar al aire dos monedas es un ejemplo de evento conjunto, pues consiste en obtener cara al lanzar al aire la primera moneda y cara al lanzar la segunda moneda.

El **complemento** del evento  $A$  (al que se le asigna el símbolo  $A'$ ) incluye todos los eventos que no son parte de  $A$ .

El complemento de una cara es una cruz, puesto que es el único evento que no es una cara. El complemento de una cara de cinco puntos es no tener una cara de cinco puntos. No obtener un lado de cinco puntos consiste en obtener un lado uno, dos, tres, cuatro o seis.

La colección de todos los eventos posibles se llama **espacio muestral**.

El espacio muestral de lanzar una moneda al aire consiste en cara y cruz. El espacio muestral cuando tiramos un dado consiste en uno, dos, tres, cuatro, cinco y seis puntos.

### EJEMPLO 4.2

#### ESPACIOS MUESTRALES Y EVENTOS

El escenario del recuadro “Uso de la estadística” en la página 122 se refiere a la empresa Consumer Electronics. La tabla 4.1 presenta los resultados de una muestra de 1,000 hogares en términos de comportamiento de compras de equipos de televisión de pantalla grande.

**TABLA 4.1**

Comportamiento de compras para equipos de televisión de pantalla grande.

#### REALMENTE LO COMPRÓ

PLANEA COMPRARLO	Sí	No	Total
<b>Sí</b>	200	50	250
<b>No</b>	100	650	750
<b>Total</b>	300	700	1,000

¿Qué es un espacio muestral? Dé ejemplos de eventos simples y eventos conjuntos.

**SOLUCIÓN** El espacio muestral consiste en las 1,000 personas encuestadas. Los eventos simples son “planea comprarlo”, “no planea comprarlo”, “compra” y “no compra”. El complemento del evento “planea comprarlo” es “no planea comprarlo”. El evento “planea comprarlo y realmente lo compra” es un evento conjunto porque quien responde debe planear comprar la televisión y realmente comprarla.

## Tablas de contingencia y diagramas de Venn

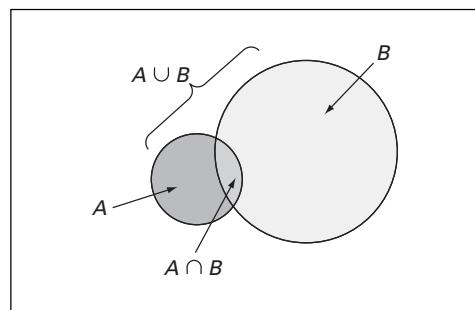
Existen diferentes formas de presentar un espacio muestral. La tabla 4.1 usa una **tabla de clasificaciones cruzadas** para presentar un espacio muestral. La tabla también se llama **tabla de contingencia** (vea la sección 2.4). Se obtienen los valores en las celdas de la tabla al subdividir el espacio muestral de los 1,000 hogares de acuerdo a si alguien planeó comprar y realmente compró un equipo de televisión de pantalla grande. Por ejemplo, 200 de quienes respondieron planearon comprar un equipo de televisión de pantalla grande y posteriormente lo compraron.

Un **diagrama de Venn** es una segunda forma de presentar un espacio muestral. Este diagrama representa gráficamente los diferentes eventos como “uniones” e “intersecciones” de círculos. La figura 4.1 presenta un diagrama de Venn típico para una situación de dos variables, en la que cada variable tiene sólo dos eventos ( $A$  y  $A'$ ,  $B$  y  $B'$ ). El círculo de la izquierda (de gris oscuro) representa todos los eventos que son parte de  $A$ . El círculo de la derecha (de gris claro) representa todos los eventos que son parte de  $B$ . El área contenida dentro del círculo  $A$  y el círculo  $B$  (área central), es la **intersección** de  $A$  y  $B$  (se escribe  $A \cap B$ ), porque es parte de  $A$  y también de  $B$ . El área total de los dos círculos es la **unión** de  $A$  y  $B$  (se escribe  $A \cup B$ ) y contiene todos los resultados que son sólo parte del evento  $A$ , sólo parte del evento  $B$ , o parte de ambos  $A$  y  $B$ . El área en el diagrama fuera de  $A \cup B$  contiene los resultados que no son parte ni de  $A$  ni de  $B$ .

Usted debe definir  $A$  y  $B$  para desarrollar un diagrama de Venn. Se puede definir a un evento como  $A$  o como  $B$ , siempre y cuando sea congruente al evaluar los diversos eventos. Para el ejemplo de Consumer Electronics, se pueden definir los eventos de la siguiente manera:

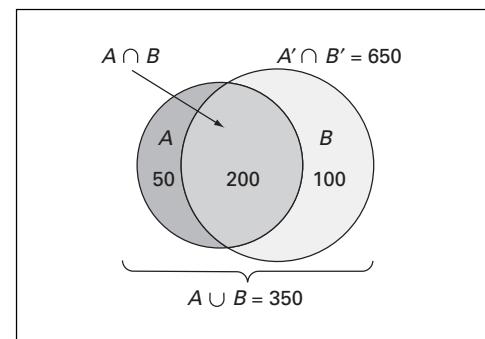
$$\begin{array}{ll} A = \text{planea comprar} & B = \text{realmente compró} \\ A' = \text{no planea comprar} & B' = \text{no lo compró} \end{array}$$

Al diseñar un diagrama de Venn (vea la figura 4.2), usted debe determinar el valor de la intersección de  $A$  y  $B$  para dividir el espacio muestral en sus partes.  $A \cap B$  consiste en los 200 hogares en los que se planeó comprar y realmente se compró el equipo de televisión de pantalla grande. El evento remanente  $A$  (planean comprar), consiste en 50 hogares en los que se planeó comprar el televisor de pantalla grande pero que finalmente no lo compraron. El remanente del evento  $B$  (realmente compraron) consiste en 100 hogares en los que no se planeó comprar un equipo de televisión de pantalla grande, pero en los que finalmente se compró uno. El resto de los 650 hogares representa a aquellos que ni planearon ni compraron un equipo de televisión de pantalla grande.



**FIGURA 4.1**

Diagrama de Venn para los eventos  $A$  y  $B$ .



**FIGURA 4.2**

Diagrama de Venn para el ejemplo de Consumer Electronics.

## Probabilidad simple (marginal)

Ahora usted está en condiciones de responder a algunas de las preguntas formuladas en el escenario “Uso de la estadística”. Como los resultados se basan en los datos recolectados en una encuesta (vea la tabla 4.1 en la página 124), puede usar el punto de vista de la probabilidad clásica empírica.

Como se estableció antes, la ley básica para las probabilidades es que varían en valor del 0 al 1. Un evento imposible tiene una probabilidad de 0 y un evento seguro de ocurrir tiene una probabilidad de 1.

**Probabilidad simple** se refiere a la probabilidad de ocurrencia de un evento simple,  $P(A)$ . En el escenario de “Uso de la estadística”, una probabilidad simple es la probabilidad de planear la compra de un equipo de televisión de pantalla grande. ¿Cómo se determina la probabilidad de seleccionar un hogar en el que se planee comprar un equipo de televisión de pantalla grande? Al utilizar la ecuación (4.1) en la página 123:

$$\text{Probabilidad de ocurrencia} = \frac{X}{T}$$

$$\begin{aligned} P(\text{planear comprar}) &= \frac{\text{número de quienes planean comprar}}{\text{número total de hogares}} \\ &= \frac{250}{1,000} = 0.25 \end{aligned}$$

Por lo tanto, hay un 0.25 (o un 25%) de probabilidad de que en un hogar se planee comprar un equipo de televisión de pantalla grande.

A la probabilidad simple también se le llama **probabilidad marginal**, porque es posible calcular el número total de los éxitos (el número total de quienes planearon comprar) a partir del margen apropiado de la tabla de contingencia (vea la tabla 4.1 en la página 124). El ejemplo 4.3 ilustra otra aplicación de la probabilidad simple.

### EJEMPLO 4.3

#### CÁLCULO DE LA PROBABILIDAD DE QUE EL EQUIPO DE TELEVISIÓN DE PANTALLA GRANDE QUE SE COMPRE SEA UN HDTV

En la encuesta de seguimiento del escenario de “Uso de la estadística”, se hicieron preguntas adicionales a 300 de los hogares en los que realmente se compró el equipo de televisión de pantalla grande. La tabla 4.2 indica las respuestas del consumidor a si el equipo comprado era un HDTV y si también compró un DVD en los últimos 12 meses.

**TABLA 4.2**

Comportamiento de compra respecto a los HDTV y los DVD.

COMPRÓ HDTV	COMPRÓ DVD		
	Sí	No	Total
<b>HDTV</b>	38	42	80
<b>No HDTV</b>	70	150	220
<b>Total</b>	108	192	300

Encuentre la probabilidad de que si en el hogar seleccionado al azar adquirieron un equipo de televisión de pantalla grande, el equipo comprado sea un HDTV.

**SOLUCIÓN** Usando las siguientes definiciones:

$$\begin{aligned} A &= \text{compró un HDTV} \\ A' &= \text{no compró un HDTV} \end{aligned}$$

$$\begin{aligned} B &= \text{compró un DVD} \\ B' &= \text{no compró un DVD} \end{aligned}$$

$$P(\text{HDTV}) = \frac{\text{número de equipos de televisión HDTV}}{\text{número total de equipos de televisión}}$$

$$= \frac{80}{300} = 0.267$$

Hay una probabilidad del 26.7% de que el equipo de televisión de pantalla grande seleccionado al azar comprado sea un HDTV.

## Probabilidad conjunta

La probabilidad marginal se refiere a la probabilidad de ocurrencia de eventos simples. La **probabilidad conjunta** se refiere a la probabilidad de ocurrencia que implica a dos o más eventos. Un ejemplo de probabilidad conjunta es la probabilidad de que se obtenga cara al lanzar la primera vez la moneda al aire y cara al lanzar por segunda vez la moneda.

En relación con la tabla 4.1 en la página 124, aquellos individuos que planearon comprar y realmente compraron el televisor de pantalla grande se identifican con los resultados de una celda singular “sí—planearon comprar y sí—realmente lo compraron”. Como el grupo está formado por 200 hogares, la probabilidad de elegir un hogar que planee comprar y realmente lo compre es

$$P(\text{planea comprar y realmente lo compra}) = \frac{\text{planea comprar y realmente compra}}{\text{número total de quienes responden}} \\ = \frac{200}{1,000} = 0.20$$

El ejemplo 4.4 también demuestra cómo se determina la probabilidad conjunta.

### EJEMPLO 4.4

#### DETERMINAR LA PROBABILIDAD CONJUNTA DE COMPRAR UN TELEVISOR DE PANTALLA GRANDE Y DE COMPRAR UN HDTV Y UN DVD

En la tabla 4.2 de la página 126, los compradores aparecen en una clasificación cruzada como HDTV o no HDTV y si en esos hogares se adquirió o no un equipo reproductor de DVD. Encuentre la probabilidad de que en los hogares seleccionados al azar, los compradores de un televisor de pantalla adquirieron un equipo HDTV y un DVD.

**SOLUCIÓN** Se emplea la ecuación (4.1) de la página 123,

$$P(\text{televisor HDTV y DVD}) = \frac{\text{número de compradores de un televisor HDTV y un DVD}}{\text{número total de compradores de un televisor de pantalla grande}} \\ = \frac{38}{300} = 0.127$$

Por lo tanto, tenemos una probabilidad del 12.7% de que el hogar seleccionado al azar en el que se adquirió un equipo de televisión de pantalla grande, haya comprado un HDTV y un DVD.

Se puede ver la probabilidad marginal de un evento en particular usando el concepto de probabilidad conjunta que se explicó antes. La probabilidad marginal de un evento consiste en un conjunto de probabilidades conjuntas. Por ejemplo, si  $B$  consiste en dos eventos,  $B_1$  y  $B_2$ , entonces  $P(A)$ , la probabilidad del evento  $A$ , consiste en la probabilidad conjunta de que el evento  $A$  ocurra con el evento  $B_1$  y la probabilidad conjunta de que el evento  $A$  ocurra con el evento  $B_2$ . Use la ecuación (4.2) para calcular las probabilidades marginales.

#### PROBABILIDAD MARGINAL

$$P(A) = P(A \text{ y } B_1) + P(A \text{ y } B_2) + \cdots + P(A \text{ y } B_k) \quad (4.2)$$

donde  $B_1, B_2, \dots, B_k$  son  $k$  eventos mutuamente excluyentes y colectivamente exhaustivos.

Los eventos mutuamente excluyentes y colectivamente exhaustivos se definen así.

Dos eventos son **mutuamente excluyentes** si ambos eventos no pueden ocurrir de manera simultánea.

Al tirar una moneda al aire, cara y cruz son eventos mutuamente excluyentes. El resultado de tirar una moneda al aire no puede ser al mismo tiempo cara y cruz.

Un conjunto de eventos es **colectivamente exhaustivo** si uno de los eventos debe ocurrir.

Cara y cruz en una moneda son eventos colectivamente exhaustivos. Uno de ellos debe ocurrir. Si no ocurre cara, entonces debe ocurrir cruz. Si cruz no ocurre, entonces debe ocurrir cara.

Ser masculino y ser femenino son eventos mutuamente excluyentes y colectivamente exhaustivos. Ninguna persona es de ambos sexos (son mutuamente excluyentes), y todos son uno u otro (son colectivamente exhaustivos).

La ecuación (4.2) sirve para calcular la probabilidad marginal de planear la compra de un equipo de televisión de pantalla grande.

$$\begin{aligned}
 P(\text{planear la compra}) &= P(\text{planear comprar y comprar}) \\
 &\quad + P(\text{planear comprar y no comprar}) \\
 &= \frac{200}{1,000} + \frac{50}{1,000} \\
 &= \frac{250}{1,000} = 0.25
 \end{aligned}$$

Usted obtendrá el mismo resultado si suma el número de resultados que conforman el evento simple “planear la compra”.

## Regla general de la adición

La regla general de la adición nos permite encontrar la probabilidad del evento “A o B”. Esta regla considera la ocurrencia de cualquiera de los eventos, evento A o evento B o ambos A y B. ¿Cómo se determina la probabilidad de que en un hogar se planee comprar o se compre realmente un equipo de televisión de pantalla grande? El evento “planear la compra o comprar realmente” incluye a todos los hogares en los que se planea comprar y todos los hogares en los que realmente se compró el equipo de televisión de pantalla grande. Revise cada celda de la tabla de contingencia (tabla 4.1, en la página 124) para determinar si es o no parte del evento. De la tabla 4.1, la celda “planea comprar y no la compró” es parte del evento porque incluye a los encuestados que planeaban comprar. La celda “no planeó comprar y realmente compró” está incluida porque contiene a los encuestados que de verdad compraron. Por último, la celda “planearon comprar y realmente compraron” tiene ambas características de interés. Por lo tanto, la probabilidad de planear comprar o realmente comprar es:

$$\begin{aligned}
 P(\text{planear comprar o realmente compró}) &= P(\text{planeó comprar y no compró realmente}) \\
 &\quad + P(\text{no planeó comprar y realmente compró}) \\
 &\quad + P(\text{planeó comprar y realmente compró}) \\
 &= \frac{50}{1,000} + \frac{100}{1,000} + \frac{200}{1,000} = \frac{350}{1,000} = 0.35
 \end{aligned}$$

A menudo encontrará más fácil determinar  $P(A \text{ o } B)$ , la probabilidad del evento A o B, mediante la **regla general de la adición** definida en la ecuación (4.3).

## REGLA GENERAL DE LA ADICIÓN

**La probabilidad de  $A$  o  $B$  es igual a la probabilidad de  $A$  más la probabilidad de  $B$  menos la probabilidad de  $A$  y  $B$ .**

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) \quad (4.3)$$

Aplicar esta ecuación al ejemplo anterior produce el siguiente resultado:

$$\begin{aligned} P(\text{planea comprar o realmente compró}) &= P(\text{planea comprar}) + P(\text{realmente compró}) \\ &\quad - P(\text{planea comprar y compró}) \\ &= \frac{250}{1,000} + \frac{300}{1,000} - \frac{200}{1,000} \\ &= \frac{350}{1,000} = 0.35 \end{aligned}$$

La regla general de la adición consiste en tomar la probabilidad de  $A$  y sumarla a la probabilidad de  $B$ , y después sustraer el evento conjunto de  $A$  y  $B$  de este total, porque el evento conjunto ya se incluyó tanto en el cálculo de la probabilidad de  $A$  y en la probabilidad de  $B$ . En relación con la tabla 4.1 en la página 124, si los resultados del evento “planean comprar” se suman a aquellos del evento “realmente compraron”, el evento conjunto “planearon comprar y realmente compraron” se incluye en cada uno de estos eventos simples. Por lo tanto, como este evento conjunto se ha contado dos veces, debe restarse para obtener el resultado correcto. El ejemplo 4.5 ilustra otra aplicación de la regla general de la adición.

**EJEMPLO 4.5****USO DE LA REGLA GENERAL DE LA ADICIÓN PARA LOS HOGARES EN LOS QUE SE COMPRARON EQUIPOS DE TELEVISIÓN DE PANTALLA GRANDE**

En el ejemplo 4.3 de la página 126, las compras se clasificaron de forma cruzada como HDTV o no HDTV y si en el hogar se compró o no un DVD. Encuentre la probabilidad de que entre los hogares en los que se adquirió un equipo de televisión de pantalla grande, se haya comprado un HDTV o un DVD.

**SOLUCIÓN** Mediante la ecuación (4.3),

$$\begin{aligned} P(\text{HDTV o DVD}) &= P(\text{HDTV}) + P(\text{DVD}) - P(\text{HDTV y DVD}) \\ &= \frac{80}{300} + \frac{108}{300} - \frac{38}{300} \\ &= \frac{150}{300} = 0.50 \end{aligned}$$

Por lo tanto, se tiene el 50.0% de probabilidad de que el hogar seleccionado al azar en el que se adquirió un equipo de televisión de pantalla grande, se haya comprado un HDTV o un DVD.

**PROBLEMAS PARA LA SECCIÓN 4.1****Aprendizaje básico**

**ASISTENCIA**  
de PH Grade

- 4.1** Se lanzan dos monedas al aire.
- Dé un ejemplo de un evento simple.
  - Dé un ejemplo de un evento conjunto.
- c.** ¿Cuál es el complemento de una cara en el primer lanzamiento?
- 4.2** Una urna contiene 12 pelotas rojas y 8 pelotas blancas. Se seleccionará una pelota de la urna.
- Dé un ejemplo de un evento simple.
  - ¿Cuál es el complemento de una pelota roja?

ASISTENCIA  
de PH Grade**4.3** A partir de la siguiente tabla de contingencia:

	<b>B</b>	<b>B'</b>
<b>A</b>	10	20
<b>A'</b>	20	40

¿Cuál es la probabilidad del

- evento  $A$ ?
- evento  $A'$ ?
- evento  $A$  y  $B$ ?
- evento  $A$  o  $B$ ?

ASISTENCIA  
de PH Grade**4.4** A partir de la siguiente tabla de contingencia:

	<b>B</b>	<b>B'</b>
<b>A</b>	10	30
<b>A'</b>	25	35

¿Cuál es la probabilidad del

- evento  $A'$ ?
- evento  $A$  y  $B$ ?
- evento  $A'$  y  $B'$ ?
- evento  $A'$  o  $B'$ ?

**Aplicación de conceptos**ASISTENCIA  
de PH Grade

**4.5** Para cada uno de los siguientes enunciados, indique si el tipo de probabilidad implicada es un ejemplo de probabilidad clásica *a priori*, probabilidad clásica empírica o probabilidad subjetiva.

- El siguiente lanzamiento de una moneda caerá cara.
- Italia ganará la Copa Mundial la próxima vez que se realice esa competencia.
- La suma de las caras de dos dados será 7.
- El tren que toma un viajero para llegar al trabajo llegará con más de 10 minutos de retraso.

**4.6** Para cada uno de los siguientes enunciados, establezca si los eventos son mutuamente excluyentes y colectivamente exhaustivos. Si no es así, renombre las categorías para hacerlos mutuamente excluyentes y colectivamente exhaustivos o explique por qué no sería útil hacerlo.

- A los votantes registrados de Estados Unidos se les preguntó si están registrados como republicanos o como demócratas.
- Quienes respondieron fueron clasificados por el tipo de automóvil que él o ella manejan: estadounidense, europeo, japonés o ninguno.
- Se les preguntó: “¿Actualmente vive en *i*) un apartamento o *ii*) en una casa?”
- Un producto fue clasificado como defectuoso o no defectuoso.

**4.7** La probabilidad de cada uno de los siguientes eventos es cero. Para cada uno de ellos establezca por qué.

- Un votante en Estados Unidos que está registrado como republicano y como demócrata.
- Un producto que es defectuoso y no defectuoso.
- Un automóvil es de marca Ford y Toyota.

✓ AUTO  
Examen

**4.8** En Estados Unidos una encuesta sobre vivienda estudió cómo llegan al trabajo los propietarios de una casa (“How People Get to Work”, *USA Today* Snapshots, 25 de febrero, 2003, 1A). Suponga que la encuesta constó de una muestra de 1,000 propietarios de casa y 1,000 inquilinos.

Maneja hacia el trabajo	Propietario de casa	Inquilino	Total
Sí	824	681	1,505
No	176	319	495
<b>Total</b>	<b>1,000</b>	<b>1,000</b>	<b>2,000</b>

- Dé un ejemplo de un evento simple.
  - Dé un ejemplo de un evento conjunto.
  - ¿Cuál es el complemento de “maneja hacia el trabajo”?
  - ¿Por qué “maneja hacia el trabajo y es un propietario de casa” es un evento conjunto?
- 4.9** En relación con la tabla de contingencia del problema 4.8, si quien responde es seleccionado al azar, ¿cuál es la probabilidad de que ella o él
- maneje hacia el trabajo?
  - maneje hacia el trabajo y sea propietario de casa?
  - maneje hacia el trabajo o sea propietario de casa?
  - Explique la diferencia en los resultados de los incisos *b* y *c*).

**4.10** Un estudio sobre el mejoramiento de la producción en una fábrica de semiconductores proporcionó datos de los defectos para una muestra de 450 placas de silicio. La siguiente tabla presenta un resumen de las respuestas a dos preguntas: “¿Se encontraron partículas en el troquel que produjo la placa?”, y “¿La placa era buena o mala?”

CONDICIÓN DEL TROQUEL			
CALIDAD DE LA PLACA	Sin partículas	Con partículas	Total
Buena	320	14	334
Mala	80	36	116
<b>Total</b>	<b>400</b>	<b>50</b>	<b>450</b>

Fuente: S. W. Hall, *Analysis of Defectivity of semiconductor Wafers by Contingency Table*, Proceedings of Institute of Environmental Sciences, vol. 1 (1994), 177-183.

- Dé un ejemplo de evento simple.
- Dé un ejemplo de evento conjunto.
- ¿Cuál es el complemento de una placa de silicio es buena?
- ¿Por qué una “placa buena” y un troquel “con partículas” es un evento conjunto?

**4.11** En relación con la tabla de contingencia del problema 4.10, si la placa de silicio se seleccionó al azar, ¿cuál es la probabilidad de que

- fuera producida con un troquel sin partículas?
- sea una placa mala y fuera producida con un troquel sin partículas?

- c. Sea una placa mala o fuera producida con un troquel con partículas?
- d. Explique la diferencia en los resultados de los incisos b) y c).

ASISTENCIA  
de PH Grade

- 4.12** ¿Es menos probable que las grandes empresas ofrezcan acciones a los miembros de su junta directiva que las empresas pequeñas o medianas? Una encuesta que realizó la Segal Company de Nueva York encontró que en una muestra de 189 empresas grandes, 40 les ofrecieron opciones de acciones a los miembros de su consejo directivo como parte de sus paquetes de compensación que no implicaban efectivo. De las empresas pequeñas y medianas, 43 de las 180 encuestadas indicaron que ofrecieron las acciones como parte de sus paquetes de compensación que no implicaban efectivo a los miembros de su consejo directivo (Kemba J. Dunham, "The Jungle: Focus on Recruitment, Pay and Getting Ahead", *The Wall Street Journal*, 21 de agosto, 2001, B6). Construya una tabla de contingencia o un diagrama de Venn para evaluar las probabilidades. Si la empresa es seleccionada al azar, ¿cuál es la probabilidad de que la compañía
- a. ofrezca opciones de acciones a los miembros de su consejo directivo?
  - b. sea una empresa pequeña o mediana y no ofrezca opciones de acciones a los miembros de su consejo directivo?
  - c. sea una empresa pequeña a mediana u ofrezca opciones de acciones a los miembros de su consejo directivo?
  - d. Explique la diferencia en los resultados de los incisos b) y c).

- 4.13** ¿Es más probable que los blancos presenten demandas por prejuicios? Una encuesta que realizó Barry Goldman ("White Fight: A Researcher Finds Whites Are More Likely to Claim Bias", *The Wall Street Journal*, Work Week, 10 de abril, 2001, A1) encontró que de 56 trabajadores despedidos, 29 presentaron demandas por prejuicios. De 407 trabajadores negros despedidos, 126 demandaron por prejuicios. Elabore una tabla de contingencia o un diagrama de Venn para evaluar las probabilidades. Si un trabajador es seleccionado al azar, ¿cuál es la probabilidad de que él o ella
- a. presente una demanda por prejuicios?

- b. sea negro y no demande por prejuicios?
- c. sea negro o demande por prejuicios?
- d. Explique la diferencia en los resultados de los incisos b) y c).

- 4.14** Una muestra de 500 personas fue seleccionada en una gran área metropolitana para estudiar el comportamiento del consumidor. Entre las preguntas estaban "¿Disfruta comprando ropa?" De 240 hombres, 136 contestaron que sí. De 260 mujeres 224 contestaron que sí. Realice una tabla de contingencia o un diagrama de Venn para evaluar las probabilidades. ¿Cuál es la probabilidad de que un encuestado elegido al azar

- a. disfrute comprando ropa?
- b. sea mujer y disfrute comprando ropa?
- c. sea mujer o disfrute comprando ropa?
- d. sea hombre o mujer?

- 4.15** Cada año se compilán las clasificaciones respecto al desempeño de los autos nuevos durante los primeros 90 días de uso. Suponga que los autos se han clasificado de acuerdo a si necesitan una garantía relacionada con reparación (sí o no) y el país en el que la empresa manufacturera tiene su sede (Estados Unidos o fuera de Estados Unidos). Con base en los datos recabados, la probabilidad de que un auto nuevo necesite de una garantía de reparación es de 0.04, la probabilidad de que el auto sea manufacturado por una empresa con sede en Estados Unidos es de 0.60, y la probabilidad de que el auto nuevo necesite una garantía de reparación y haya sido manufacturado por una empresa con sede en Estados Unidos es de 0.025. Elabore una tabla de contingencia o un diagrama de Venn para evaluar la probabilidad de una garantía relacionada con la reparación. ¿Cuál es la probabilidad de que un auto nuevo seleccionado al azar

- a. necesite una garantía relacionada con reparaciones?
- b. necesite una garantía relacionada con reparaciones y sea manufacturado por una empresa con sede en estados Unidos?
- c. necesite una garantía de reparación o fue manufacturado por una empresa con sede en Estados Unidos?
- d. necesite una garantía de reparación o no fue manufacturado por una empresa con sede en Estados Unidos?

## 4.2 PROBABILIDAD CONDICIONAL

### Cálculo de probabilidades condicionales

Cada ejemplo en la sección 4.1 implicó encontrar la probabilidad de un evento muestreado del espacio muestral completo. ¿Cómo se determina la probabilidad de un evento si cierta información acerca de los eventos implicados es ya conocida?

La **probabilidad condicional** se refiere a la probabilidad del evento *A*, dada información acerca de la ocurrencia de otro evento *B*.

#### PROBABILIDAD CONDICIONAL

La probabilidad de *A* dado *B* es igual a la probabilidad de *A* y *B* dividida por la probabilidad de *B*

$$P(A | B) = \frac{P(A \text{ y } B)}{P(B)} \quad (4.4a)$$

La probabilidad de  $B$  dado  $A$  es igual a la probabilidad de  $A$  y  $B$  dividida por la probabilidad de  $A$

$$P(B | A) = \frac{P(A \text{ y } B)}{P(A)} \quad (4.4b)$$

donde

$P(A \text{ y } B)$  = probabilidad conjunta de  $A$  y  $B$

$P(A)$  = probabilidad marginal de  $A$

$P(B)$  = probabilidad marginal de  $B$

En relación con el escenario de “Uso de la estadística” que se refiere a la compra de un equipo de televisión de pantalla grande, suponga que en cierto hogar se planea comprar un equipo de televisión de pantalla grande. Ahora, ¿cuál es la probabilidad de que en ese hogar se compre realmente el equipo de televisión? En este ejemplo el objetivo es encontrar  $P(\text{compra real} | \text{planea comprar})$ . Aquí se le proporciona la información de que el hogar planea comprar el equipo de televisión de pantalla grande. Por lo tanto, el espacio muestral no consiste en todos los 1,000 hogares de la encuesta. Consiste sólo en aquellos que realmente compraron el equipo de televisión de pantalla grande. De 250 de esos hogares, 200 compraron realmente el equipo de televisión de pantalla grande. Por lo tanto (vea la tabla 4.1 en la página 124 o la figura 4.2 en la página 125), la probabilidad de que en un hogar realmente se compre un equipo de televisión de pantalla grande dado que lo planeó comprar es

$$P(\text{realmente compró} | \text{planeó comprar}) = \frac{\text{planeó comprar y realmente compró}}{\text{planeó comprar}}$$

$$= \frac{200}{250} = 0.80$$

También es posible usar la ecuación (4.4b) para calcular este resultado.

$$P(B | A) = \frac{P(A \text{ y } B)}{P(A)}$$

donde

evento  $A$  = planeó comprar

evento  $B$  = realmente compró

Entonces

$$P(\text{realmente compró} | \text{planeó comprar}) = \frac{200/1,000}{250/1,000}$$

$$= \frac{200}{250} = 0.80$$

El ejemplo 4.6 ilustra aún más la probabilidad condicional.

#### EJEMPLO 4.6

ENCONTRAR LA PROBABILIDAD CODICIONAL REFERENTE A LOS HOGARES QUE REALMENTE COMPRARON UN EQUIPO DE TELEVISIÓN DE PANTALLA GRANDE

La tabla 4.3 en la página 126 es una tabla de contingencia que se refiere a si el hogar compró un HDTV y un DVD. De los hogares que compraron un HDTV, ¿cuál es la probabilidad de que también hayan comprado un DVD?

**SOLUCIÓN** Como se sabe en qué hogares compraron un HDTV, el espacio muestral se reduce a 80 hogares. De estos 80, 38 también compraron un DVD. Por lo tanto, la probabilidad de que un hogar comprara un DVD, dado que el hogar compró un HDTV es:

$$P(\text{compró DVD} \mid \text{compró HDTV}) = \frac{\text{número que compró HDTV y DVD}}{\text{número que compró HDTV}}$$

$$= \frac{38}{80} = 0.475$$

Si se usa la ecuación (4.4a) de la página 131:

$$A = \text{compró DVD} \quad B = \text{compró HDTV}$$

entonces

$$P(A \mid B) = \frac{P(A \text{ y } B)}{P(B)} = \frac{38/300}{80/300} = 0.475$$

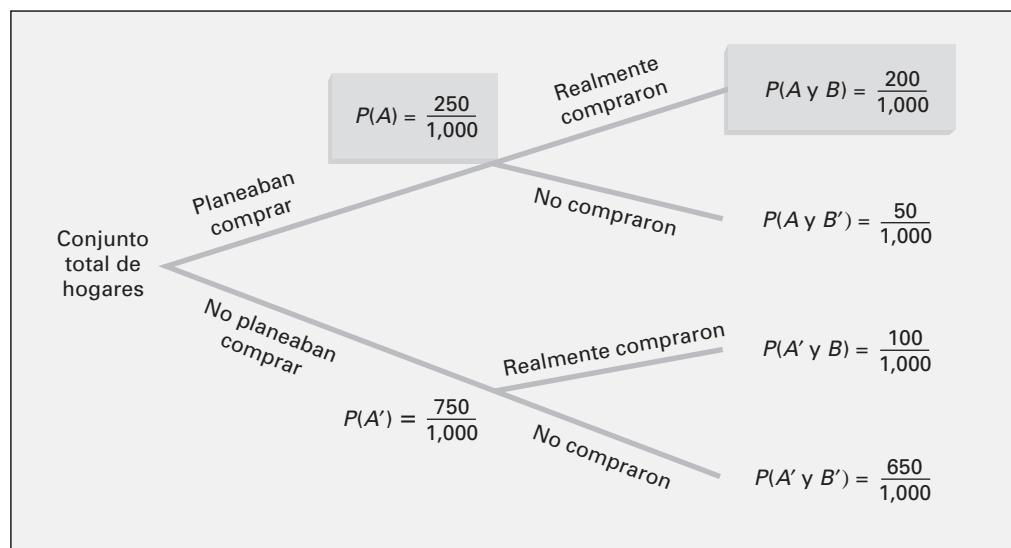
Entonces, dado que en un hogar se compró un HDTV, hay un 47.5% de posibilidades de que también haya comprado un DVD. Se puede comparar esta probabilidad condicional con la probabilidad marginal de comprar un DVD, la cual es de  $108/300 = 0.36$ , o del 36%. Estos resultados indican que los hogares en los que se compró un HDTV tienen más probabilidades de comprar un DVD que los hogares que compraron un equipo de televisión de pantalla grande que no es un HDTV.

## Árboles de decisión

En la tabla 4.1 en la página 124, los hogares se clasifican de acuerdo con sus planes de comprar o de si realmente compraron el equipo de televisión de pantalla grande. Un **árbol de decisión** es una alternativa para la tabla de contingencia. La figura 4.3 representa el árbol de decisión para este ejemplo.

**FIGURA 4.3**

Árbol de decisión para el ejemplo de Consumer Electronics.



En la figura 4.3 iniciando a la izquierda con el conjunto total de hogares, se abren dos “ramas” para indicar si planearon o no comprar el equipo de televisión de pantalla grande. Cada una de estas ramas tiene dos subramas, correspondientes a si el hogar realmente compró o no el equipo de televisión de pantalla grande. Las probabilidades al final de las ramas iniciales representan la probabilidad

marginal de  $A$  y  $A'$ . La probabilidad al final de cada una de las cuatro subramas representa la probabilidad conjunta de cada combinación de eventos  $A$  y  $B$ . Calcule la probabilidad condicional dividiendo la probabilidad conjunta por la probabilidad marginal apropiada.

Por ejemplo, para calcular la probabilidad de que cierto hogar realmente compró luego de que planeó comprar el equipo de televisión de pantalla grande, tome  $P(\text{planeó comprar y realmente compró})$  y divídalo por  $P(\text{planeó comprar})$ . A partir de la figura 4.3

$$\begin{aligned} P(\text{realmente compró} \mid \text{planeó comprar}) &= \frac{200/1,000}{250/1,000} \\ &= \frac{200}{250} = 0.80 \end{aligned}$$

El ejemplo 4.7 ilustra cómo construir un árbol de decisión.

### EJEMPLO 4.7

#### CREACIÓN DEL ÁRBOL DE DECISIÓN PARA LOS HOGARES EN LOS QUE SE COMPRÓ UN EQUIPO DE TELEVISIÓN DE PANTALLA GRANDE

Utilice los datos cruzados clasificados de la tabla 4.2 en la página 126 y realice el árbol de decisión. Use el árbol de decisión para encontrar la probabilidad de que en un hogar se haya comprado un DVD, dado que el hogar compró un HDTV.

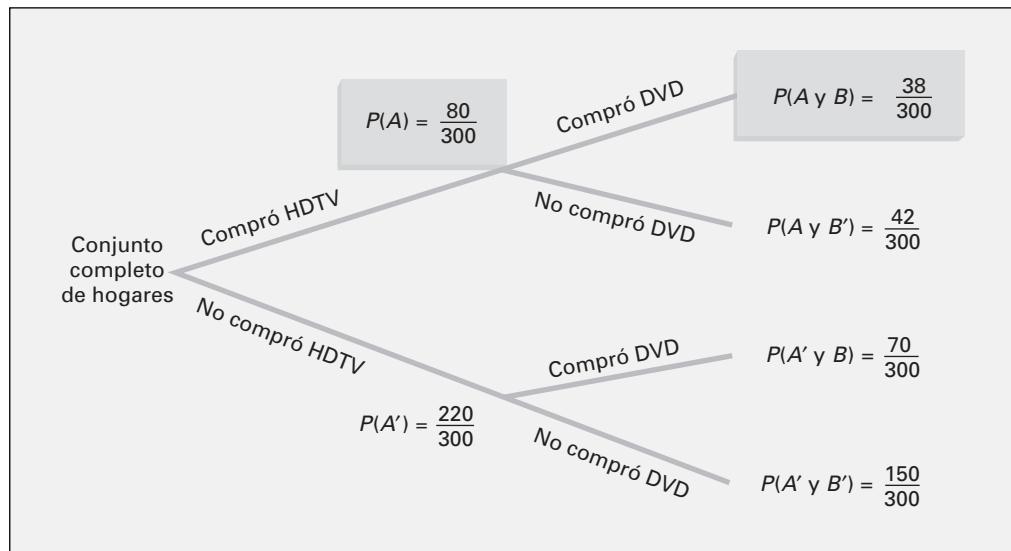
**SOLUCIÓN** El árbol de decisión para haber comprado un DVD y un HDTV se muestra en la figura 4.4. Emplee la ecuación (4.4b) en la página 132 y las siguientes definiciones:

$$A = \text{compró HDTV} \quad B = \text{compró un DVD}$$

$$P(B \mid A) = \frac{P(A \text{ y } B)}{P(A)} = \frac{38/300}{80/300} = 0.475$$

**FIGURA 4.4**

Árbol de decisión para haber comprado un DVD y un HDTV.



### Independencia estadística

En el ejemplo relacionado con la compra de equipos de televisión de pantalla grande, la probabilidad condicional es de  $200/250 = 0.80$  de que en el hogar seleccionado realmente se haya comprado un equipo de televisión de pantalla grande, luego de que se planeó comprarlo. La probabilidad sim-

ple de seleccionar un hogar que realmente hizo la compra es de  $300/1,000 = 0.30$ . Estos resultados muestran que el conocimiento previo de que en el hogar se planeó comprar afectó la probabilidad de que el hogar realmente comprara el equipo de televisión. En otras palabras, el resultado de un evento es *dependiente* del resultado de un segundo evento.

Cuando el resultado de un evento *no* afecta la probabilidad de ocurrencia de otro evento, se dice que los eventos son estadísticamente independientes. La **independencia estadística** se determina mediante la ecuación (4.5).

### INDEPENDENCIA ESTADÍSTICA

Dos eventos  $A$  y  $B$  son estadísticamente independientes si y sólo si

$$P(A | B) = P(A) \quad (4.5)$$

donde

$P(A | B)$  = probabilidad condicional de  $A$  dado  $B$

$P(A)$  = probabilidad marginal de  $A$

El ejemplo 4.8 demuestra el uso de la ecuación (4.5).

### EJEMPLO 4.8

### DETERMINACIÓN DE LA INDEPENDENCIA ESTADÍSTICA

En el estudio de seguimiento de 300 hogares que realmente compraron un equipo de televisión de pantalla grande, se preguntó a los encuestados si estaban satisfechos con sus compras. La tabla 4.3 de respuestas cruzadas clasifica las respuestas a la pregunta sobre la satisfacción dependiendo de si el equipo de televisión era un HDTV.

**TABLA 4.3**

Satisfacción con la compra de equipos de televisión de pantalla grande.

TIPO DE TELEVISIÓN	¿SATISFECHO CON LA COMPRA?		
	Sí	No	Total
HDTV	64	16	80
No HDTV	176	44	220
Total	240	60	300

Determine si estar satisfecho con la compra y el tipo de televisor comprado son estadísticamente independientes.

**SOLUCIÓN** Para estos datos,

$$P(\text{satisfecho} | \text{HDTV}) = \frac{64/300}{80/300} = \frac{64}{80} = 0.80$$

lo que es igual a

$$P(\text{satisfecho}) = \frac{240}{300} = 0.80$$

Así que estar satisfecho con la compra y el tipo de equipo de televisión comprado son estadísticamente independientes. El conocimiento de un evento no afecta la probabilidad del otro evento.

## Reglas de multiplicación

Al manipular la fórmula de la probabilidad condicional, es posible determinar la probabilidad conjunta  $P(A \text{ y } B)$  de la probabilidad condicional de un evento. La **regla general de la multiplicación** se obtiene con la ayuda de la ecuación (4.4a) en la página 131.

$$P(A | B) = \frac{P(A \text{ y } B)}{P(B)}$$

y se resuelve la probabilidad conjunta  $P(A \text{ y } B)$ .

### REGLA GENERAL DE LA MULTIPLICACIÓN

La probabilidad de  $A$  y  $B$  es igual a la probabilidad de  $A$  dado  $B$  por la probabilidad de  $B$ .

$$P(A \text{ y } B) = P(A | B)P(B) \quad (4.6)$$

El ejemplo 4.9 demuestra el uso de la regla general de la multiplicación.

### EJEMPLO 4.9

#### USO DE LA REGLA DE MULTIPLICACIÓN

Considere los 80 hogares en los que se compró un HDTV. En la tabla 4.3 en la página 135 se observa que en 64 hogares están satisfechos con su compra y en 16 hogares no están satisfechos. Suponga que se seleccionan al azar dos hogares de los 80 que realizaron la compra. Encuentre la probabilidad de que ambos hogares estén satisfechos con su adquisición.

**SOLUCIÓN** Aquí se emplea la regla de la siguiente manera. Si:

$$A = \text{segundo hogar seleccionado está satisfecho}$$

$$B = \text{primer hogar seleccionado está satisfecho}$$

entonces, mediante la ecuación (4.6)

$$P(A \text{ y } B) = P(A | B)P(B)$$

La probabilidad de que el primer hogar esté satisfecho con la compra es de 64/80. Sin embargo, la probabilidad de que el segundo hogar también esté satisfecho con la compra depende del resultado de la primera selección. Si el primer hogar no se devuelve a la muestra después de determinar el nivel de satisfacción (muestreo sin sustitución), entonces el número de hogares restantes será de 79. Si el primer hogar está satisfecho, la probabilidad de que el segundo hogar también esté satisfecho es 63/79, porque en la muestra permanecen 63 hogares satisfechos. Por lo tanto,

$$P(A \text{ y } B) = \left(\frac{63}{79}\right)\left(\frac{64}{80}\right) = 0.6380$$

Hay 63.80% de posibilidades de que ambos hogares muestreados estén satisfechos con sus compras.

La **regla de multiplicación para eventos independientes** se obtiene al sustituir  $P(A)$  por  $P(A | B)$  en la ecuación (4.6).

### REGLA DE MULTIPLICACIÓN PARA EVENTOS INDEPENDIENTES

Si  $A$  y  $B$  son estadísticamente independientes, la probabilidad de  $A$  y  $B$  es igual a la probabilidad de  $A$  por la probabilidad de  $B$ .

$$P(A \text{ y } B) = P(A)P(B) \quad (4.7)$$

Si esta regla sirve para dos eventos,  $A$  y  $B$ , entonces  $A$  y  $B$  son estadísticamente independientes. Por lo tanto, existen dos maneras de determinar la independencia estadística.

1. Los eventos  $A$  y  $B$  son estadísticamente independientes si y sólo si  $P(A | B) = P(A)$ .
2. Los eventos  $A$  y  $B$  son estadísticamente independientes si y sólo si  $P(A \text{ y } B) = P(A)P(B)$ .

## Probabilidad marginal usando la regla general de la multiplicación

En la sección 4.1 se definió la probabilidad marginal con la ayuda de la ecuación (4.2) de la página 127. Es factible expresar la fórmula para la probabilidad marginal mediante la regla general de la multiplicación. Si

$$P(A) = P(A \text{ y } B_1) + P(A \text{ y } B_2) + \cdots + P(A \text{ y } B_k)$$

entonces, al usar la regla general de la multiplicación, la ecuación (4.8) define la probabilidad marginal.

### PROBABILIDAD MARGINAL USANDO LA REGLA GENERAL DE LA MULTIPLICACIÓN

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \cdots + P(A | B_k)P(B_k) \quad (4.8)$$

donde  $B_1, B_2, \dots, B_k$  son los eventos  $k$  mutuamente excluyentes y colectivamente exhaustivos.

Para ilustrar esta ecuación, consulte la tabla 4.1 en la página 124. Al utilizar la ecuación (4.8), la probabilidad de planear una compra es:

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2)$$

donde

$P(A)$  = probabilidad de “planea comprar”

$P(B_1)$  = probabilidad de “realmente compró”

$P(B_2)$  = probabilidad de “no compró”

$$\begin{aligned} P(A) &= \left( \frac{200}{300} \right) \left( \frac{300}{1,000} \right) + \left( \frac{50}{700} \right) \left( \frac{700}{1,000} \right) \\ &= \frac{200}{1,000} + \frac{50}{1,000} = \frac{250}{1,000} = 0.25 \end{aligned}$$

## PROBLEMAS PARA LA SECCIÓN 4.2

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 4.16** A partir de la siguiente tabla de contingencia:

	<b><i>B</i></b>	<b><i>B'</i></b>
<i>A</i>	10	20
<i>A'</i>	20	40

¿Cuál es la probabilidad de

- a.  $A | B$ ?
- b.  $A | B'$ ?
- c.  $A' | B'$ ?
- d. Los eventos  $A$  y  $B$  son estadísticamente independientes?

- 4.17** A partir de la siguiente tabla de contingencia:

	<b><i>B</i></b>	<b><i>B'</i></b>
<i>A</i>	10	30
<i>A'</i>	25	35

¿Cuál es la probabilidad de

- a.  $A | B$ ?
- b.  $A' | B'$ ?
- c.  $A | B'$ ?
- d. Los eventos  $A$  y  $B$  son estadísticamente independientes?

ASISTENCIA  
de PH Grade

- 4.18** Si  $P(A \text{ y } B) = 0.4$  y  $P(B) = 0.8$ , encuentre  $P(A | B)$ .

ASISTENCIA  
de PH Grade

- 4.19** Si  $P(A) = 0.7$  y  $P(B) = 0.6$ , y si  $A$  y  $B$  son estadísticamente independientes, encuentre  $P(A \text{ y } B)$ .

ASISTENCIA  
de PH Grade

- 4.20** Si  $P(A) = 0.3$  y  $P(B) = 0.4$ , y si  $P(A \text{ y } B) = 0.2$ , ¿ $A$  y  $B$  son estadísticamente independientes?

### Aplicación de conceptos

AUTO  
Examen

- 4.21** En Estados Unidos una encuesta sobre vivienda estudió cómo llegan al trabajo los propietarios de casa ("How People Get to Work," *USA Today* Snapshots, 25 de febrero, 2003, 1A). Suponga que la encuesta consistió en una muestra de 1,000 propietarios de casa y 1,000 inquilinos.

Maneja al trabajo	Propietario	Inquilino	Total
Sí	824	681	1,505
No	176	319	495
Total	1,000	1,000	2,000

- Si una persona responde que maneja hacia su trabajo, ¿cuál es la probabilidad de que él o ella sea propietario de su casa?
- Si quien responde es un propietario, ¿cuál es la probabilidad de que él o ella maneje hacia el trabajo?
- Explique la diferencia en los resultados de los incisos *a* y *b*.
- ¿Son estos dos eventos, manejar hacia el trabajo y si quien responde renta la casa o es propietario de ella, estadísticamente independientes?

- 4.22** Un estudio de mejoramiento de la producción de un fabricante de semiconductores proporcionó datos de defectos para una muestra de 450 placas de silicio. La siguiente tabla presenta un resumen de las respuestas a dos preguntas: "¿Se encontraron partículas en el troquel que produjo la placa de silicio?", y "¿La placa era buena o mala?"

CONDICIÓN DEL TROQUEL			
CALIDAD DE LA PLACA	Sin partículas	Partículas	Total
Buena	320	14	334
Mala	80	36	116
Total	400	50	450

Fuente: S.W. Hall, *Analysis of Defectivity of Semiconductor Wafers by Contingency Table*, Proceedings Institute of Environmental Sciences, Vol. 1 (1994), 177-183.

- Suponga que sabe que una placa de silicio es mala. ¿Cuál es la probabilidad de que fuera producida con un troquel que tenía partículas?
- Suponga que sabe que una placa de silicio es buena. ¿Cuál es la probabilidad de que fuera producida con un troquel que tenía partículas?
- ¿Son estos dos eventos, una placa buena y un troquel sin partículas, estadísticamente independientes? Explique su respuesta.

ASISTENCIA  
de PH Grade

- 4.23** ¿Es menos probable que las grandes empresas ofrezcan acciones a los miembros de su consejo directivo que las empresas pequeñas o medianas? Una encuesta que realizó Segal Company de Nueva York encontró que en una muestra de 189 empresas grandes, 40 ofrecieron opciones de acciones a los miembros de su consejo directivo como parte de sus paquetes de compensación sin efectivo. Entre las compañías pequeñas y medianas, 43 de las 180 encuestadas indicaron que ofrecieron las opciones de acciones como parte de sus paquetes de compensación que no implicaban efectivo a los miembros de su consejo directivo (Kemba J. Dunham, "The Jungle: Focus on Recruitment, Pay and Getting Ahead", *The Wall Street Journal*, 21 de agosto, 2001, B6).

- Si una empresa es grande, ¿cuál es la probabilidad de que ofrezca opciones de acciones a los miembros de su consejo directivo?
- Si la empresa es de pequeña a mediana, ¿cuál es la probabilidad de que ofrezca opciones de acciones a los miembros de su consejo directivo?
- ¿Es el tamaño de la empresa estadísticamente independiente de si se ofrecen opciones de acciones a los miembros del consejo directivo? Explique su respuesta.

- 4.24** ¿Es más probable que los blancos presenten demandas por prejuicios? Una encuesta conducida por Barry Goldman ("White Fight: A Researcher Finds Whites Are More Likely to Claim Bias", *The Wall Street Journal*, Work Week, 10 de abril, 2001, A1) encontró que de 56 trabajadores despedidos, 29 presentaron demandas por prejuicios. De 407 trabajadores negros despedidos, 126 demandaron por prejuicios.

- Si un trabajador es blanco, ¿cuál es la probabilidad de que haya presentado una demanda por prejuicios?
- Si un trabajador presentó una demanda por prejuicios, ¿cuál es la probabilidad de que sea blanco?
- Explique las diferencias en los resultados de los incisos *a* y *b*.
- ¿Son los dos eventos "ser blanco" y "demandar por prejuicios" estadísticamente independientes? Explique su respuesta.

- 4.25** Se seleccionó una muestra de 500 encuestados en un área metropolitana grande para estudiar el comportamiento del consumidor. Los resultados fueron los siguientes:

GÉNERO			
DISFRUTA COMPRANDO	Masculino	Femenino	Total
Sí	136	224	360
No	104	36	140
Total	240	260	500

- Suponga que el encuestado elegido es mujer. ¿Cuál es la probabilidad de que ella no disfrute de comprar ropa?
- Suponga que el encuestado elegido disfruta de comprar ropa. ¿Cuál es la probabilidad de que el individuo sea hombre?
- ¿Disfrutar comprando ropa y el género del individuo son eventos estadísticamente independientes? Explique su respuesta.

**4.26** Cada año se compilan las clasificaciones respecto al desempeño de los autos nuevos durante los primeros 90 días de uso. Suponga que los autos se han clasificado de acuerdo a si el auto necesita una garantía relacionada con reparación (sí o no) y el país en el que la empresa manufacturera tiene su sede (Estados Unidos o fuera de Estados Unidos). Con base en los datos recabados, la probabilidad de que un auto nuevo necesite de una garantía de reparación es de 0.04, la probabilidad de que el auto sea manufacturado por una empresa con sede en Estados Unidos es de 0.60, y la probabilidad de que el auto nuevo necesite una garantía de reparación y haya sido manufacturado por una empresa con sede en Estados Unidos es de 0.025.

- Suponga que usted sabe que una empresa asentada en los Estados Unidos manufacturó el auto. ¿Cuál es la probabilidad de que el auto necesite una garantía de reparación?
- Suponga que sabe que una empresa con sede en Estados Unidos no manufacturó el auto. ¿Cuál es la probabilidad de que el auto necesite una garantía de reparación?
- ¿La necesidad de una garantía de reparación y la sede de la empresa que manufactura autos son estadísticamente independientes?

**4.27** En 34 de los 54 años que van desde 1950 a 2003, S&P 500 terminó muy alto después de los primeros cinco días de transacciones. En 29 de esos 34 años S&P terminó alto durante el año. ¿Iniciar bien la primera semana es buen presagio para el año que vendrá? La siguiente tabla muestra el desempeño de la primera semana y el desempeño anual en este periodo de 54 años.

DESEMPEÑO ANUAL DE S&P 500

PRIMERA SEMANA	Alto	Bajo
Alto	29	5
Bajo	10	10

Fuente: Adaptado de Aaron Luchetti, "Stocks Enjoy a Good First Week", The Wall Street Journal, 12 de enero, 2004, C1.

- Si el año se selecciona al azar, ¿cuál es la probabilidad de que S&P haya terminado alto ese año?

**b.** Si S&P 500 terminó más alto después de los primeros cinco días de transacciones, ¿cuál es la probabilidad de que haya terminado alto en el año?

**c.** ¿Los dos eventos, desempeño en la primera semana y desempeño anual, son estadísticamente independientes? Explique su respuesta.

**d.** En 2004, S&P 500 estuvo arriba un 0.9% después de los primeros cinco días. Vea el desempeño anual del 2004 de S&P 500 en [finance.yahoo.com](http://finance.yahoo.com). Comente los resultados.

**4.28** Se usa un mazo de cartas tradicional para organizar un juego. Hay cuatro palos (corazones, diamantes, tréboles y espadas), cada uno tiene 13 cartas (as, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, reina y rey), que hacen un total de 52 cartas. Este mazo completo se baraja perfectamente y usted recibirá dos cartas del mazo sin reemplazo.

- ¿Cuál es la probabilidad de que ambas cartas sean reinas?
- ¿Cuál es la probabilidad de que la primera carta sea un 10 y la segunda carta sea un 5 o un 6?
- Si la muestra fuera con reemplazo, ¿cuál sería la respuesta al inciso a)?
- En el juego de blackjack, las cartas con figura (jack, reina, rey) valen 10 puntos y el as vale 1 u 11 puntos. Todas las demás cartas cuentan por su valor. Usted logra hacer blackjack si sus dos cartas totalizan 21 puntos. ¿Cuál es la probabilidad de obtener un blackjack en este problema?

**4.29** Una caja de nueve guantes de golf contiene dos guantes para la mano izquierda y siete guantes para la mano derecha.

- a.** Si dos guantes de la caja se seleccionan al azar sin reemplazo, ¿cuál es la probabilidad de que ambos guantes seleccionados sean para la mano derecha?
- b.** Si dos guantes de la caja se seleccionan al azar sin reemplazo, ¿cuál es la probabilidad de que uno de los guantes seleccionados sea para la mano derecha y uno para la mano izquierda?
- c.** Si se seleccionan con reemplazo tres guantes, ¿cuál es la probabilidad de que los tres guantes sean para la mano izquierda?
- d.** Si la muestra fuera con reemplazo, ¿cuáles serían las respuestas para los incisos a) y b)?

### 4.3 TEOREMA DE BAYES

El **teorema de Bayes** se utiliza para revisar probabilidades previamente calculadas cuando se posee nueva información. Desarrollado por el Reverendo Thomas Bayes en el siglo XVIII (vea la referencia 1), el teorema de Bayes es una extensión de lo que ha aprendido hasta ahora acerca de la probabilidad condicional.

El teorema de Bayes se aplica a la siguiente situación. La empresa Consumer Electronics está considerando comercializar un nuevo modelo de televisor. En el pasado, el 40% de los equipos de televisión que la empresa lanzó al mercado tuvieron éxito y el 60% no fueron exitosos. Antes de lanzar al mercado el equipo de televisión, el departamento de investigación de mercados realiza un extenso estudio y entrega un reporte, ya sea favorable o desfavorable. En el pasado, el 80% de los equipos de televisión exitosos habían recibido un reporte de investigación favorable y el 30% de los equipos de televisión no exitosos habían recibido un reporte de investigación favorable. Para los nuevos modelos de televisión bajo consideración, el departamento de investigación de mercado ha entregado un reporte favorable. ¿Cuál es la probabilidad de que el equipo de televisión tenga éxito en el mercado?

ASISTENCIA  
de PH Grade

El teorema de Bayes se desarrolla a partir de la definición de la probabilidad condicional. Para encontrar la probabilidad condicional de  $B$  dado  $A$ , considere la ecuación (4.4b) [que se presentó en la página 132 y que de nuevo se reproduce]:

$$P(B|A) = \frac{P(A \text{ y } B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

El teorema de Bayes se obtiene por sustitución de la ecuación (4.8) en la página 137 para  $P(A)$  en la ecuación de arriba.

### TEOREMA DE BAYES

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_k)P(B_k)} \quad (4.9)$$

donde  $B_i$  es el  $i$ -ésimo evento de los  $k$  eventos mutuamente excluyentes y colectivamente exhaustivos.

Para usar la ecuación (4.9) para el ejemplo del televisor comercializado, sea

$$\begin{array}{ll} \text{evento } S = \text{equipo de televisión exitoso} & \text{evento } F = \text{reporte favorable} \\ \text{evento } S' = \text{equipo de televisión no exitoso} & \text{evento } F' = \text{reporte desfavorable} \end{array}$$

y

$$\begin{array}{ll} P(S) = 0.40 & P(F|S) = 0.80 \\ P(S') = 0.60 & P(F|S') = 0.30 \end{array}$$

Entonces, al utilizar la ecuación (4.9),

$$\begin{aligned} P(S|F) &= \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S')P(S')} \\ &= \frac{(0.80)(0.40)}{(0.80)(0.40) + (0.30)(0.60)} \\ &= \frac{0.32}{0.32 + 0.18} = \frac{0.32}{0.50} \\ &= 0.64 \end{aligned}$$

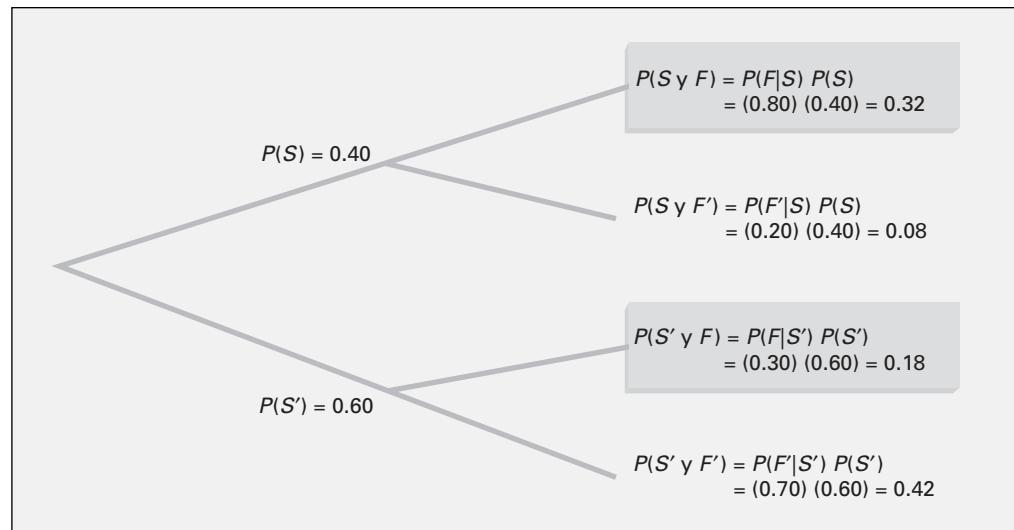
La probabilidad de un equipo de televisión exitoso, dado que se recibió un reporte favorable, es de 0.64. Así pues, la probabilidad de un equipo de televisión no exitoso, dado que se recibió un reporte favorable, es de  $1 - 0.64 = 0.36$ . La tabla 4.4 resume el cálculo de las probabilidades y la figura 4.5 presenta el árbol de decisión.

**TABLA 4.4** Cálculo del teorema de Bayes para el ejemplo de la comercialización del televisor

Evento $S_i$	Probabilidad previa $P(S_i)$	Probabilidad condicional $P(F S_i)$	Probabilidad conjunta $P(F S_i)P(S_i)$	Probabilidad revisada $P(S_i F)$
$S = \text{equipo de televisión exitoso}$	0.40	0.80	0.32	$0.32/0.50 = 0.64 = P(S F)$
$S' = \text{equipo de televisión no exitoso}$	0.60	0.30	$\frac{0.18}{0.50}$	$0.18/0.50 = 0.36 = P(S' F)$

## FIGURA 4.5

## Árbol de decisión para la comercialización del nuevo equipo de televisión.



El ejemplo 4.10 aplica el teorema de Bayes a un problema de diagnóstico médico.

### EJEMPLO 4.10

USO DEL TEOREMA DE BAYES EN UN PROBLEMA DE DIAGNÓSTICO MÉDICO

La probabilidad de que una persona tenga una determinada enfermedad es de 0.03. Existen pruebas de diagnóstico médico disponibles para determinar si una persona tiene realmente la enfermedad. Si la enfermedad realmente está presente, la probabilidad de que la prueba de diagnóstico médico dé un resultado positivo (indicando la presencia de la enfermedad) es de 0.90. Si la enfermedad no está presente, la probabilidad de obtener un resultado positivo (indicando la presencia de la enfermedad) es de 0.02. Suponga que la prueba de diagnóstico médico dio un resultado positivo (indicando la presencia de la enfermedad). ¿Cuál es la probabilidad de que la enfermedad esté realmente presente? ¿Cuál es la probabilidad de un resultado positivo?

## SOLUCIÓN

Sea	evento $D$ = tiene la enfermedad	evento $T$ = la prueba es positiva
y	evento $D'$ = no tiene la enfermedad	evento $T'$ = la prueba es negativa

$$\begin{array}{ll} P(D) = 0.03 & P(T | D) = 0.90 \\ P(D') = 0.97 & P(T | D') = 0.02 \end{array}$$

Al emplear la ecuación (4.9) de la página 140,

$$\begin{aligned}
 P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D')P(D')} \\
 &= \frac{(0.90)(0.03)}{(0.90)(0.03) + (0.02)(0.97)} \\
 &= \frac{0.0270}{0.0270 + 0.0194} = \frac{0.0270}{0.0464} \\
 &= 0.582
 \end{aligned}$$

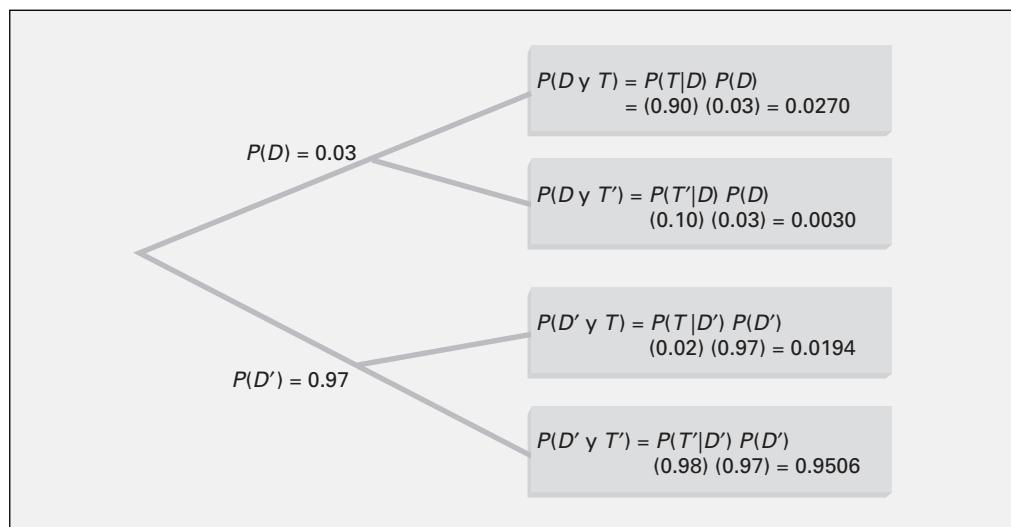
La probabilidad de que la enfermedad esté realmente presente dado que un resultado positivo ha ocurrido (indicando la presencia de la enfermedad) es de 0.582. La tabla 4.5 resume el cálculo de las probabilidades y la figura 4.6 presenta el árbol de decisión.

**TABLA 4.5** Cálculos del teorema de Bayes para el problema de diagnóstico médico

Evento $D_i$	Probabilidad previa $P(D_i)$	Probabilidad condicional $P(T   D_i)$	Probabilidad conjunta $P(T   D_i)P(D_i)$	Probabilidad revisada $P(D_i   T)$
$D = \text{tiene la enfermedad}$	0.03	0.90	0.0270	$0.0270/0.0464 = 0.582 = P(D   T)$
$D' = \text{no tiene la enfermedad}$	0.97	0.02	$\frac{0.0194}{0.0464}$	$0.0194/0.0464 = 0.418 = P(D'   T)$

**FIGURA 4.6**

Árbol de decisión para el problema de diagnóstico médico.



El denominador en el teorema de Bayes representa  $P(T)$ , la probabilidad de un resultado positivo en la prueba, el cual en este caso es de 0.0464 o un 4.64%.

## PROBLEMAS PARA LA SECCIÓN 4.3

### Aprendizaje básico

**4.30** Si  $P(B) = 0.05$ ,  $P(A | B) = 0.80$ ,  $P(B') = 0.95$ , y  $P(A | B') = 0.40$ , encuentre  $P(B | A)$ .

**4.31** Si  $P(B) = 0.30$ ,  $P(A | B) = 0.60$ ,  $P(B') = 0.70$ , y  $P(A | B') = 0.50$ , encuentre  $P(B | A)$ .

### Aplicación de conceptos

**4.32** En el ejemplo 4.10 de la página 141, suponga que la probabilidad de que la prueba de diagnóstico médico dé un resultado positivo si la enfermedad no está presente se reduce de 0.02 a 0.01. A partir de esta información,

- Si la prueba de diagnóstico médico ha resultado positiva (indicando la presencia de la enfermedad), ¿cuál es la probabilidad de que la enfermedad esté realmente presente?
- Si la prueba de diagnóstico médico ha dado un resultado negativo (indicando que la enfermedad no está presente), ¿cuál es la probabilidad de que la enfermedad no esté presente?

ASISTENCIA  
de PH Grade ✓ AUTO  
Examen

**4.33** Un directivo de publicidad estudia los hábitos de ver televisión de hombres y mujeres casados durante las horas de mayor audiencia. Con base en los registros anteriores, el directivo ha determinado que durante las horas de mayor audiencia los maridos ven televisión el 60% del tiempo. Cuando el marido ve televisión, la esposa también lo hace el 40% del tiempo. Cuando el marido no ve televisión, la esposa ve televisión el 30% del tiempo. Encuentre la probabilidad de que

- si la esposa ve televisión, el esposo también lo haga.
- la esposa vea televisión durante las horas de mayor audiencia.

ASISTENCIA  
de PH Grade

**4.34** La empresa Olive Construction está determinando si debería presentar una oferta para un nuevo centro comercial. En el pasado, el principal competidor de Olive, la empresa Base Construction, ha propuesto ofertas el 70% del tiempo. Si Base Construction no presenta ofertas para un trabajo, la probabilidad de que Olive Construction ob-

tenga el trabajo es de 0.50. Si Base Construction propone una oferta para el trabajo, la probabilidad de que Olive Construction obtenga el trabajo es de 0.25.

- Si la empresa Olive Construction obtiene el trabajo, ¿cuál es la probabilidad de que la empresa Base Construction no haya propuesto una oferta?
- ¿Cuál es la probabilidad de que la empresa Olive Construction obtenga el trabajo?

**4.35** Los trabajadores despedidos que se volvieron empresarios porque no encontraron empleo en otra empresa se conocen como *empresarios por necesidad*. El *Wall Street Journal* reporta que estos empresarios tienen menos posibilidad de crecimiento en los grandes negocios que los *empresarios por elección* (Jeff Bailey, “Desire—More Than Need—Builds a Business”, *The Wall Street Journal*, 21 de mayo, 2001, B4). Este artículo establece que el 89% de los empresarios en Estados Unidos lo son por elección y que el 11% son empresarios por necesidad. Sólo el 2% de los empresarios por necesidad esperan que su nuevo negocio dé empleo a 20 o más personas dentro de los siguientes cinco años, mientras que el 14% de los empresarios por elección esperan emplear por lo menos a 20 personas dentro de los siguientes cinco años.

- Si se selecciona al azar a un empresario y éste espera que su nuevo negocio emplee a 20 o más personas dentro de los siguientes cinco años, ¿cuál es la probabilidad de que este individuo sea un empresario por elección?
- Discuta las posibles razones por las que los empresarios por elección tienen más posibilidades de creer que sus negocios crecerán.

**4.36** El editor de una empresa editorial de libros de texto está tratando de decidir si publicar un libro de texto propuesto de es-

tadística en los negocios. Información sobre los libros de texto previamente publicados indica que el 10% tiene un enorme éxito, el 20% tiene un éxito moderado, el 40% ni gana ni pierde y el 30% fracasa. Sin embargo, antes de tomar la decisión de publicar, el libro se revisa. En el pasado, el 99% de los libros exitosos recibieron revisiones favorables, el 70% de los de éxito moderado recibieron revisiones favorables, el 40% de los que ni ganaron ni perdieron recibieron revisiones favorables, y el 20% de los que fracasaron recibieron revisiones favorables.

- Si el texto propuesto recibió revisiones favorables, ¿cómo debe revisar el editor las probabilidades de los diferentes resultados para tomar en cuenta esta información?
- ¿Qué proporción de libros de texto reciben revisiones favorables?

**4.37** Un servicio municipal de títulos tiene tres categorías de clasificación (*A*, *B* y *C*). Suponga que el año pasado, de los títulos municipales distribuidos a lo largo de Estados Unidos, el 70% entró en la categoría *A*, el 20% entró en la categoría *B*, y el 10% se consideró como *C*. De los títulos municipales clasificados en *A*, el 50% se distribuyó en ciudades, el 40% en suburbios y el 10% en áreas rurales. De los títulos municipales clasificados como *B*, el 60% se distribuyó en ciudades, el 20% en suburbios y el 20% en áreas rurales. De los títulos municipales clasificados como *C*, el 90% se distribuyó en ciudades, el 5% en suburbios y el 5% en áreas rurales.

- Si un nuevo título municipal va a distribuirse en una ciudad, ¿cuál es la probabilidad de que reciba una clasificación *A*?
- ¿Qué proporción de títulos municipales se distribuye en ciudades?
- ¿Qué proporción de títulos municipales se distribuye en suburbios?

## 4.4 REGLAS DE CONTEO

En la ecuación (4.1) de la página 123, la probabilidad de ocurrencia de un resultado se definió como el número de formas en las que el resultado ocurre, dividido por el número total de resultados posibles. En muchos casos, hay un gran número de posibles resultados y es difícil determinar el número exacto. En estas circunstancias se han desarrollado las reglas para contar el número posible de resultados. En este apartado se presentan cinco diferentes reglas de conteo.

### REGLA DE CONTEO 1

Si cualquiera de los eventos  $k$  mutuamente excluyentes y colectivamente exhaustivos pueden ocurrir en cada uno de los ensayos  $n$ , el número de posibles resultados es igual a

$$k^n \quad (4.10)$$

### EJEMPLO 4.11

### REGLA DE CONTEO 1

Suponga que se lanza una moneda al aire cinco veces. ¿Cuál es el número de diferentes resultados posibles (la secuencia de caras y cruces)?

**SOLUCIÓN** Si se lanza al aire una moneda (que tiene dos lados) cinco veces, con la ecuación (4.10), el número de resultados es  $2^5 = 2 \times 2 \times 2 \times 2 \times 2 = 32$ .

**EJEMPLO 4.12****LANZAR UN DADO DOS VECES**

Suponga que se lanza un dado dos veces. ¿Cuántos posibles resultados diferentes pueden ocurrir?

**SOLUCIÓN** Si un dado (con seis caras) se lanza dos veces, con la ecuación (4.10), el número de resultados diferentes es  $6^2 = 36$ .

La segunda regla de conteo es una versión más general que la primera y permite al número posible de eventos de diferir de ensayo a ensayo.

**REGLA DE CONTEO 2**

Si hay  $k_1$  eventos en el primer ensayo,  $k_2$  eventos en el segundo ensayo, ... y  $k_n$  eventos en el  $n$ -ésimo ensayo, entonces el número posible de resultados es

$$(k_1)(k_2) \dots (k_n) \quad (4.11)$$

**EJEMPLO 4.13****REGLA DE CONTEO 2**

Un departamento estatal de vehículos automotores desea saber cuántos números para las placas están disponibles si las placas incluyen tres letras seguidas por tres números.

**SOLUCIÓN** Mediante la ecuación (4.11), se sabe que si una placa incluye tres letras seguidas por tres números (0 hasta 9), el número total de resultados posibles es  $(26)(26)(26)(10)(10)(10) = 17,576,000$ .

**EJEMPLO 4.14****DETERMINAR EL NÚMERO DE DIFERENTES CENAS**

El menú de un restaurante tiene un precio fijo para las cenas completas que consisten en un aperitivo, un platillo principal, una bebida y un postre. Se tiene la posibilidad de elegir entre cinco aperitivos, 10 platillos, tres bebidas y seis postres. Determine el número total de cenas posibles.

**SOLUCIÓN** Mediante la ecuación (4.11), se sabe que el número total de cenas posibles es  $(5)(10)(3)(6) = 900$ .

La tercera regla de conteo está relacionada con el cálculo del número de maneras en las que un conjunto de cosas puede arreglarse en orden.

**REGLA DE CONTEO 3**

El número de maneras en el que las  $n$  cosas pueden arreglarse en orden es

$$n! = (n)(n - 1) \dots (1) \quad (4.12)$$

donde  $n!$  se llama el *factorial de  $n$*  y  $0!$  se define como 1.

**EJEMPLO 4.15****REGLA DE CONTEO 3**

Si un conjunto de seis libros de texto se colocan en una repisa, ¿de cuántas formas es posible ordenar estos seis libros de texto?

**SOLUCIÓN** Para empezar, hay que precisar que cualquiera de los seis libros podría ocupar la primera posición en la repisa. Una vez que se llene la primera posición, hay cinco libros a elegir para

llenar la segunda. Se continúa con este procedimiento de asignación hasta que todas las posiciones estén ocupadas. El número de formas en las que es posible acomodar los seis libros es

$$n! = 6! = (6)(5)(4)(3)(2)(1) = 720$$

En muchos casos se necesita saber el número de formas en las que un subconjunto de un grupo completo de cosas puede arreglarse en *orden*. Cada posible arreglo es llamado permutación.

#### REGLA DE CONTEO 4

*Permutaciones:* El número de maneras para arreglar  $X$  objetos seleccionados de  $n$  objetos en orden es

$${}_n P_X = \frac{n!}{(n - X)!} \quad (4.13)$$

#### EJEMPLO 4.16

#### REGLA DE CONTEO 4

Si se modifica el ejemplo 4.15, y entonces tenemos seis libros de texto pero sólo hay espacio para cuatro libros en la repisa, ¿de cuántas maneras es posible acomodar estos libros en la repisa?

**SOLUCIÓN** Con la ecuación (4.13), se sabe que el número de arreglos ordenados de cuatro libros seleccionados de seis libros es igual a

$${}_n P_X = \frac{n!}{(n - X)!} = \frac{6!}{(6 - 4)!} = \frac{(6)(5)(4)(3)(2)(1)}{(2)(1)} = 360$$

En muchas situaciones no interesa el *orden* de los resultados, sino sólo el número de maneras en las que  $X$  objetos pueden seleccionarse a partir de  $n$  cosas, *sin consideración de orden*. Esta regla se llama la regla de las **combinaciones**.

#### REGLA DE CONTEO 5

*Combinaciones:* El número de maneras de seleccionar  $X$  objetos a partir de  $n$  objetos, sin considerar el orden, es igual a

$${}_n C_X = \frac{n!}{X!(n - X)!} \quad (4.14)$$

Al comparar esta regla con la anterior, se observa que difiere sólo en la inclusión del término  $X!$  en el denominador. Cuando se usan las permutaciones, todos los arreglos de los  $X$  objetos son distinguibles. Con las combinaciones, los  $X!$  posibles arreglos de objetos son irrelevantes.

#### EJEMPLO 4.17

#### REGLA DE CONTEO 5

Se modifican el ejemplo 4.16, de manera que el orden de los libros en la repisa sea irrelevante, ¿de cuántas maneras es posible arreglar estos libros en la repisa?

**SOLUCIÓN** Al utilizar la ecuación (4.14), se sabe que el número de combinaciones de cuatro libros seleccionados de seis libros es igual a

$${}_n C_X = \frac{n!}{X!(n - X)!} = \frac{6!}{4!(6 - 4)!} = \frac{(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)(2)(1)} = 15$$

## PROBLEMAS PARA LA SECCIÓN 4.4

### Aplicación de conceptos

 **AUTO Examen** **4.38** Si hay 10 preguntas de opción múltiple en un examen, cada una con tres posibles respuestas, ¿cuántas diferentes secuencias de respuestas hay?

**4.39** La cerradura de la bóveda de un banco consta de tres discos, cada uno con 30 posiciones. Para que la bóveda abra, cada uno de los tres discos debe de estar en la posición correcta.

- ¿Cuántas posibles “combinaciones de disco” diferentes hay para esta cerradura?
- ¿Cuál es la probabilidad de que, si se selecciona al azar la posición de cada disco, se abra la bóveda del banco?
- Explique por qué las “combinaciones de disco” no son combinaciones matemáticas que se expresen con la ecuación (4.14).

**4.40 a.** Si se lanza al aire una moneda siete veces, ¿cuántos resultados diferentes son posibles?

- Si se lanza un dado siete veces, ¿cuántos resultados diferentes son posibles?
- Discuta las diferencias en sus respuestas a los incisos *a* y *b*.

 **AUTO Examen** **4.41** Una marca de jeans para dama está disponible en siete tallas, tres diferentes colores y tres diferentes estilos. ¿Cuántos diferentes pares de jeans debe ordenar el administrador para tener un par de cada tipo?

**4.42** A usted le gustaría hacer una ensalada que contenga lechuga, jitomate, pepino y germinados. Se dirige hacia el supermercado con el fin de comprar un tipo de cada uno de estos ingredientes. Ahí descubre que existen ocho tipos de lechuga, cuatro tipos de jitomates, tres tipos de pepinos y tres tipos de germinados en venta. ¿Cuántos tipos diferentes de ensaladas tiene para elegir?

 **AUTO Examen** **4.43** Si cada letra se utiliza una vez, ¿cuántas “palabras” diferentes de cuatro letras es posible formar con las letras E, L, O y V?

**4.44** En la Liga Mayor de Béisbol hay cinco equipos en la División Occidental de la Liga Nacional: Arizona, Los Ángeles, San Francisco, San Diego y Colorado. ¿Cuántos diferentes órdenes de terminar hay para estos cinco equipos? ¿Piensa que todos estos órdenes son igualmente posibles? Discuta su respuesta.

**4.45** En relación con el problema 4.44, ¿cuántos diferentes órdenes de terminar son posibles para las primeras cuatro posiciones?

**4.46** Un jardinero tiene seis filas disponibles en su jardín para colocar tomates, berenjena, pimientos, pepinos, frijoles y lechuga. A cada vegetal le asignará una y sólo una fila. ¿Cuántas formas hay de ordenar los vegetales en este jardín?

 **AUTO Examen** **4.47** El gran premio Big Triple, en la pista de carreras local, consiste en elegir el orden correcto de llegada para los tres primeros caballos de la novena carrera. Si hay 12 caballos inscritos en la novena carrera de hoy, ¿cuántos resultados Big Triple habrá?

**4.48** La Quiniela, en la pista local de carreras, consiste en elegir los caballos que llegarán en primero y segundo lugares en una carrera *sin consideración* del orden. Si ocho caballos están inscritos en la carrera, ¿cuántas combinaciones de quiniela habrá?

 **AUTO Examen** **4.49** Un estudiante tiene siete libros que quisiera colocar en un estuche. Sin embargo, en él sólo caben cuatro libros. Sin importar el orden, ¿cuántas formas hay de colocar los cuatro libros en el estuche?

**4.50** En una lotería diaria los dos números ganadores se seleccionan de entre 100 números. ¿Cuántas posibles combinaciones de números ganadores son posibles?

**4.51** Una lista de lecturas para un curso contiene 20 artículos. Cuántas formas diferentes hay para elegir tres artículos de esta lista?



## 4.5 CONSIDERACIONES ÉTICAS Y PROBABILIDAD

Los problemas éticos surgen cuando cualquier afirmación acerca de la probabilidad se presenta ante el público, particularmente cuando estas afirmaciones forman parte de una campaña de publicidad para un producto o servicio. Por desgracia, mucha gente no se siente a gusto con los conceptos numéricos (vea la referencia 3 al final del capítulo) y tiende a malinterpretar el significado de la probabilidad. En algunos casos, la mala interpretación no es intencional, pero en otros, la publicidad trata de engañar al cliente potencial, lo que constituye una falta de ética.

Un ejemplo de la aplicación potencialmente no ética de la publicidad se relaciona con la publicidad para las loterías estatales. Cuando una persona compra un billete de lotería, selecciona un conjunto de números (por ejemplo, 6) de una lista mayor de números (como 54). Aunque virtualmente todos los participantes saben que tienen muy pocas posibilidades de ganar la lotería, también tienen muy poca idea de qué tan improbable es que ellos seleccionen los 6 números ganadores de la lista de 54 números. Tienen una idea aún mucho menor de la probabilidad de ganar un premio de consolación si acierran a 4 o 5 de los números ganadores.

Con estos antecedentes, podría detenerse a considerar como engañoso y posiblemente no ético el reciente comercial de una lotería estatal que afirmaba: "No pararemos hasta que hagamos de cada uno un millonario". Como la lotería aporta millones de dólares a la tesorería estatal, el Estado nunca dejará de permitirla, aunque en la vida nadie podrá estar seguro de volverse millonario al ganar la lotería.

Otro ejemplo de una aplicación potencialmente no ética de la publicidad se relaciona con una carta de inversión que promete un 90% de probabilidad de obtener una ganancia del 20% anual sobre la inversión. Para que el reclamo de la carta fuera ético, el servicio de inversión necesita *a*) explicar la base de esta estimación de probabilidad, *b*) hacer la afirmación de probabilidad en otro formato, tal como 9 oportunidades en 10, y *c*) explicar qué sucede a la inversión en el 10% de los casos en los que no se logra una ganancia del 20% (por ejemplo, ¿se pierde la inversión?).

## PROBLEMAS PARA LA SECCIÓN 4.5

### Aplicación de conceptos

**4.52** Escriba un mensaje publicitario para la lotería estatal que describa de forma ética la probabilidad de ganar.

**4.53** Escriba un mensaje publicitario para la carta de inversión, que afirme éticamente la probabilidad de obtener una ganancia del 20%.

## RESUMEN

En este capítulo se desarrollaron conceptos concernientes a la probabilidad básica, la probabilidad condicional, el teorema de Bayes y las reglas de conteo. En el siguiente capítulo se desa-

rrollarán importantes distribuciones de probabilidad discreta, como la distribución binomial y la distribución de Poisson.

## FÓRMULAS IMPORTANTES

### Probabilidad de ocurrencia

$$\text{Probabilidad de ocurrencia} = \frac{X}{T} \quad (4.1)$$

### Probabilidad marginal

$$P(A) = P(A \text{ y } B_1) + P(A \text{ y } B_2) + \cdots + P(A \text{ y } B_k) \quad (4.2)$$

### Regla general de la adición

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) \quad (4.3)$$

### Probabilidad condicional

$$P(A | B) = \frac{P(A \text{ y } B)}{P(B)} \quad (4.4a)$$

$$P(B | A) = \frac{P(A \text{ y } B)}{P(A)} \quad (4.4b)$$

### Independencia estadística

$$P(A | B) = P(A) \quad (4.5)$$

### Regla general de la multiplicación

$$P(A \text{ y } B) = P(A | B)P(B) \quad (4.6)$$

### Regla de la multiplicación para eventos independientes

$$P(A \text{ y } B) = P(A)P(B) \quad (4.7)$$

### Probabilidad marginal usando la regla general de la multiplicación

$$P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \cdots + P(A | B_k)P(B_k) \quad (4.8)$$

### Teorema de Bayes

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \cdots + P(A | B_k)P(B_k)} \quad (4.9)$$

### Regla de conteo 1

$$k^n \quad (4.10)$$

### Regla de conteo 2

$$(k_1)(k_2) \dots (k_n) \quad (4.11)$$

### Factoriales

$$n! = (n)(n-1) \dots (1) \quad (4.12)$$

### Permutaciones

$${}_n P_X = \frac{n!}{(n-X)!} \quad (4.13)$$

### Combinaciones

$${}_n C_X = \frac{n!}{X!(n-X)!} \quad (4.14)$$

## CONCEPTOS CLAVE

Árbol de decisión 133	Independencia estadística 135	Probabilidad subjetiva 123
Colectivamente exhaustivo 128	Intersección 125	Regla de multiplicación para eventos independientes 136
Combinaciones 149	Mutuamente excluyentes 128	Regla general de la adición 128
Complemento 124	Permutación 145	Regla general de la multiplicación 136
Diagrama de Venn 125	Probabilidad 122	Tabla de clasificaciones cruzadas 125
Espacio muestral 124	Probabilidad clásica <i>a priori</i> 123	Tabla de contingencia 125
Evento 124	Probabilidad clásica empírica 123	Teorema de Bayes 139
Evento conjunto 124	Probabilidad condicional 131	Unión 125
Evento imposible 123	Probabilidad conjunta 127	
Evento seguro 123	Probabilidad marginal 126	
Evento simple 124	Probabilidad simple 126	

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**4.54** ¿Cuáles son las diferencias entre la probabilidad clásica *a priori*, la probabilidad clásica empírica y la probabilidad subjetiva?

**4.55** ¿Cuál es la diferencia entre un evento simple y un evento conjunto?

**4.56** ¿Cómo se utiliza la regla de la adición para encontrar la probabilidad de ocurrencia de un evento *A o B*?

**4.57** ¿Cuál es la diferencia entre los eventos mutuamente excluyentes y los eventos colectivamente exhaustivos?

**4.58** ¿Cómo es que se relaciona la probabilidad condicional con el concepto de independencia estadística?

**4.59** ¿Cómo difiere la regla de la multiplicación para eventos que son independientes y para aquellos que no lo son?

**4.60** ¿Cómo se utiliza el teorema de Bayes para revisar las probabilidades a la luz de una nueva información?

**4.61** ¿Cuál es la diferencia entre una permutación y una combinación?

### Aplicación de conceptos

**4.62** Una compañía embotelladora de bebidas refrescantes mantiene registros concernientes al número de botellas inaceptables de las máquinas de llenado y sellado. Con base en datos anteriores, la probabilidad de que una botella provenga de la máquina I y sea no aceptable es de 0.01, y la probabilidad de que una botella provenga de la máquina II y sea no aceptable es de 0.025. La mitad de las botellas se llenan en la máquina I y la otra mitad se llena en la máquina II. Si se selecciona al azar una botella, ¿cuál es la probabilidad de que

- sea una botella no aceptable?
- haya sido llenada en la máquina I y sea una botella aceptable?

- haya sido llenada en la máquina I o sea una botella aceptable?
- Suponga que sabe que la botella fue llenada en la máquina I. ¿Cuál es la probabilidad de que sea no aceptable?
- Suponga que sabe que la botella es no aceptable. ¿Cuál es la probabilidad de que haya sido llenada en la máquina I?
- Explique la diferencia en las respuestas de los incisos *d* y *e*.

(Sugerencia: Realice una tabla de contingencia  $2 \times 2$  o un diagrama de Venn para evaluar las probabilidades.)

**4.63** Una encuesta preguntó a trabajadores qué aspectos de su actividad laboral eran extremadamente importantes. Los resultados en porcentajes son los siguientes:

¿Es el aspecto excesivamente importante?		
Aspecto del trabajo	Hombres	Mujeres
<b>Buena relación con el jefe</b>	63%	77%
<b>Equipo actualizado</b>	59	69
<b>Recursos para hacer el trabajo</b>	55	74
<b>Facilidad para llegar</b>	48	60
<b>Horario flexible</b>	40	53
<b>Posibilidad de trabajar en casa</b>	21	34

Fuente: "Snapshot", USA Today, 15 de mayo, 2000.

Suponga que la encuesta se basa en las respuestas de 500 hombres y 500 mujeres. Elabore una tabla de contingencia para las diferentes respuestas concernientes a cada aspecto del trabajo. Si quien responde es elegido al azar, ¿cuál es la probabilidad de que

- él o ella sientan que tener una buena relación con el jefe es un aspecto importante del trabajo?
- él o ella sientan que llegar fácilmente es un aspecto importante del trabajo?
- la persona sea hombre y sienta que una buena relación con el jefe es un aspecto importante del trabajo?

- d. la persona sea mujer y sienta que tener flexibilidad en las horas es un aspecto importante del trabajo?
- e. Dado que la persona siente que tener una buena relación con el jefe es un aspecto importante del trabajo, ¿cuál es la probabilidad de que la persona sea hombre?
- f. ¿Las cosas que los trabajadores dicen que son extremadamente importantes del trabajo son estadísticamente independientes del género de quien responde? Explique por qué.

**4.64** Muchas empresas usan sitios Web para llevar a cabo transacciones de negocios, tales como tomar órdenes o realizar intercambios financieros. Estos sitios se llaman sitios Web de transacciones públicas. Un análisis de 490 firmas listadas en el Fortune 500 identifica a las firmas con base en su nivel de ventas y en función de si la firma tiene o no un sitio Web de transacciones públicas (D. Young, y J. Benamati, "A Cross-Industry Analysis of Large Firm Transactional Public Web Sites", *Mid American Journal of Business*, 19(2994), 37-46). Los resultados de este análisis se muestran en la siguiente tabla.

SITIO WEB DE TRANSACCIONES PÚBLICAS		
VENTAS (EN MILLONES DE DÓLARES)	Sí	No
Mayores a \$10 mil millones	71	88
Hasta \$10 mil millones	99	232

- a. Dé un ejemplo de un evento simple y de un evento conjunto.
- b. ¿Cuál es la probabilidad de que una firma en el Fortune 500 tenga un sitio Web de transacciones públicas?
- c. ¿Cuál es la probabilidad de que una firma en el Fortune 500 tenga ventas que superen los 10,000 millones de dólares y un sitio Web de transacciones públicas?
- d. ¿Los eventos ventas que superan los 10,000 millones de dólares y tiene un sitio Web de transacciones públicas son independientes? Explique por qué.

**4.65** El dueño de un restaurante que sirve platillos estilo Continental, está interesado en estudiar los patrones de órdenes de los clientes para el periodo de fin de semana que va de viernes a domingo. Se llevaron registros que indican la demanda de postre durante el mismo periodo de tiempo. El dueño decidió estudiar otras dos variables junto con si ordenan o no un postre: el género del individuo y si ordenó o no el platillo de res. Los resultados son los siguientes:

GÉNERO			
ORDENÓ POSTRE	Masculino	Femenino	Total
Sí	96	40	136
No	224	240	464
Total	320	280	600

PLATILLO DE RES			
ORDENÓ POSTRE	Sí	No	Total
Sí	71	65	136
No	116	348	464
Total	187	413	600

Un mesero se aproxima a una mesa para tomar la orden. ¿Cuál es la probabilidad de que el primer cliente que ordene en la mesa

- a. ordene un postre?
- b. ordene un postre o una entrada de res?
- c. sea mujer y no ordene postre?
- d. sea mujer o no ordene postre?
- e. Suponga que la primera persona a la que el mesero toma la orden de postre es mujer. ¿Cuál es la probabilidad de que ella no ordene postre?
- f. ¿El género y ordenar postre son eventos estadísticamente independientes?
- g. ¿Ordenar un platillo de res es un evento estadísticamente independiente de si la persona ordena o no postre?

**4.66** Correos electrónicos que contienen publicidad comercial no solicitada, llamados spam, son borrados de forma rutinaria por el 80% de los usuarios antes de leerlos. Además, un pequeño porcentaje de quienes leen los spam continúan con el proceso y compran artículos. Sin embargo, muchas empresas usan esta publicidad no solicitada porque el costo es extremadamente bajo. Movies Unlimited es una empresa de Filadelfia que trabaja en la venta de videos y DVD vía correo electrónico, y es una de las que tienen más éxito generando ventas a través de esta forma de mercadeo. Ed Weiss, gerente general de Movies Unlimited, estima que entre el 15 y 20% de sus receptores de correos electrónicos leen la publicidad. Más aún, aproximadamente el 15% de quienes leen la publicidad hacen un pedido (Stacy Forster, "E-Marketers Look to Polish Spam's Rusty Image", *The Wall Street Journal*, 22 de mayo, 2002, D2).

- a. De acuerdo con la estimación más baja de Weiss referente a que la probabilidad de que un receptor lea la publicidad es de 0.15, ¿cuál es la probabilidad de que el receptor lea la publicidad y haga un pedido?
- b. La empresa Movies Unlimited usa una base de datos de 175,000 clientes para enviar publicidad por correo electrónico. Si se envía publicidad por esta vía a cada cliente de la base de datos, ¿cuántos clientes se espera que lean la publicidad y hagan un pedido?
- c. Si la probabilidad de que un receptor lea la publicidad es de 0.20, ¿cuál es la probabilidad de que el receptor lea la publicidad y haga un pedido?
- d. ¿Cuál sería su respuesta al inciso b) si se supone que la probabilidad de que un receptor lea la publicidad es de 0.20?

ASISTENCIA  
de PH Grade

**4.67** En febrero de 2002, el peso argentino perdió el 70% de su valor en relación con el dólar de Estados Unidos. Esta devaluación incrementó drásticamente el precio de los productos de importación. De acuerdo con una encuesta conducida por AC Nielsen en abril de 2002, el 68% de los clientes en Argentina compraban menos productos que antes de la devaluación, el 24% compraba el mismo número de productos y el 8% compraba más productos. Además, en una tendencia hacia la compra de marcas menos caras, el 88% indicó que habían cambiado las marcas que compraban. (Michelle Wallin, "Argentines Hone Art of Shopping in a Crisis", *The Wall Street Journal*, 28 de mayo, 2002, A15.) Suponga que se reportó el siguiente conjunto de resultados.

MARCAS COMPRADAS	NÚMERO DE PRODUCTOS COMPRADOS			
	Menos	Igual	Más	Total
La misma	10	14	24	48
Cambió	262	82	8	352
Total	272	96	32	400

¿Cuál es la probabilidad de que un cliente seleccionado al azar:

- compre menos productos que antes?
- compre el mismo número o más productos que antes?
- compre menos productos y cambie de marca?
- Si el cliente cambió las marcas que compraba, ¿cuál es la probabilidad de que combre menos productos que antes?
- Compare los resultados del inciso *a*) con los del *d*).

**ASISTENCIA de PH Grade** **4.68** Los vehículos utilitarios deportivos (SUV), camionetas y pick-ups son generalmente más propensos a volcarse que otros autos. En 1997, el 24.0% de los accidentes en carretera implicaron una volcadura; en el 15.8% de todos los accidentes en ese año estuvieron implicadas SUV, camionetas y pick-ups. Cuando los accidentes no implicaban volcaduras, en el 5.6% de todos ellos participaron SUV, camionetas y pick-ups (Anna Wilde Mathews, "Ford Ranger, Chevy Tracker Tilt in Test", *The Wall Street Journal*, 14 de julio, 1999, A2). Considera las siguientes definiciones:

*A* = Una SUV, camioneta o pick-up que participan en accidentes.

*B* = Un accidente que implicó una volcadura.

- Utilice el teorema de Bayes para encontrar la probabilidad de que el accidente haya implicado una volcadura, dado que participó una SUV, camioneta o pick-up.
- Compare los resultados del inciso *a*) con la probabilidad de que el accidente haya implicado una volcadura, y comente si las SUV, camionetas y pick-ups son más propensas a sufrir accidentes por volcadura.

**ASISTENCIA de PH Grade** **4.69** La prueba de ELISA (prueba de enzimas ligadas al inmunosorbente es el tipo más común de prueba de monitoreo para detectar el VIH. Un resultado positivo de una prueba ELISA indica la presencia del VIH. La prueba de ELISA tiene un alto grado de sensibilidad (para detectar la infección) y de especificidad (para detectar la no infección). (Visite el sitio HIVInsite, en [HIVInsite.ucsf.edu/](http://HIVInsite.ucsf.edu/).) Suponga que la probabilidad de que una persona esté infectada con el VIH para una determinada población es de 0.015. Si el VIH está realmente presente, la probabilidad de que la prueba de ELISA dé un resultado positivo es de 0.995. Si el VIH no está realmente presente, la probabilidad de un resultado positivo en la prueba de ELISA es de 0.01. Use el teorema de Bayes para encontrar la probabilidad de que el VIH esté realmente presente si la prueba de ELISA da un resultado positivo.

## CASO WEB

Aplique sus conocimientos acerca de las tablas de contingencia y de la aplicación apropiada de las probabilidades simple y conjunta en la continuación del Caso Web del capítulo 3.

Visite la página Web StockTout Guaranteed Investment Package en [www.prenhall.com/Springville/ST\\_Guaranteed.htm](http://www.prenhall.com/Springville/ST_Guaranteed.htm). Lea las quejas y examine los datos que las apoyan. Después responda a las siguientes preguntas:

- ¿Qué tan precisa es la afirmación de la probabilidad de éxito para el StockTout's Guaranteed Investment Packa-

ge? ¿De qué formas es engañosa esta afirmación? ¿Cómo podría calcular y establecer la probabilidad de tener una ganancia anual no menor al 15%?

- ¿Qué error se cometió al reportar la afirmación del 7% de probabilidad? Usando la tabla que se encuentra en la página Web "Winning Probabilities" [ST\\_Guaranteed3.htm](http://ST_Guaranteed3.htm), calcule la probabilidad adecuada para el grupo de inversionistas.
- ¿Existe algún tipo de cálculo de probabilidades que sea apropiado para el nivel en el servicio de inversión? ¿Por qué?

## REFERENCIAS

- Kirk, R. L. (ed.), *Statistical Issues: A Reader for the Behavioral Sciences* (Belmont, CA: Wadsworth, 1972).
- Microsoft Excel 2003* (Redmond, WA: Microsoft Corp., 2002).
- Paulos, J. A., *Innumeracy* (Nueva York: Hill and Wang, 1988).

# Apéndice 4 Uso del software para la probabilidad básica

## A4.1 EXCEL

### Para probabilidades básicas

Abra el archivo **Probabilidades.xls**. Esta hoja de trabajo ya contiene las entradas para la tabla 4.2 de la página 126. Para adaptar esta hoja de trabajo a otros problemas, cambie las entradas en las celdas sombreadas de las filas 3 hasta la 6.

*O* si está usando el PHStat2, seleccione **PHStat → Probabilidad y Distribuciones de probabilidad → Probabilidades simple y conjunta** para generar una hoja de trabajo en la que pueda ingresar sus datos de probabilidad en las celdas vacías y sombreadas de las filas 3 hasta la 6.

### Para el teorema de Bayes

Abra el archivo **Bayes.xls**. Esta hoja de trabajo ya contiene las entradas para la tabla 4.4 de la página 140. Para adaptar esta hoja de trabajo a otros problemas, cambie las entradas para las probabilidades previa y condicional en la celda sombreada clasificada B5:C6.



# CAPÍTULO 5

## Algunas importantes distribuciones de probabilidad discreta

---

**USO DE LA ESTADÍSTICA:** El sistema de información contable de la empresa de remodelaciones Saxon

### 5.1 DISTRIBUCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA

Valor esperado de una variable aleatoria discreta  
Varianza y desviación estándar de una variable aleatoria discreta

### 5.2 DISTRIBUCIÓN BINOMIAL

### 5.3 DISTRIBUCIÓN DE POISSON

### A.5 USO DE SOFTWARE PARA DISTRIBUCIONES DE PROBABILIDAD DISCRETAS

*A5.1 Excel*

*A5.2 Minitab*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- Las propiedades de una distribución de probabilidad
- A calcular el valor esperado, la varianza y la desviación estándar de una distribución de probabilidad
- A calcular las probabilidades de las distribuciones binomial y de Poisson
- Cómo utilizar las distribuciones binomial y de Poisson para resolver problemas financieros

# USO DE LA ESTADÍSTICA



## El sistema de información contable de la empresa de remodelaciones Saxon

Además de recabar, procesar, almacenar y transformar la información financiera, los sistemas de información contable la distribuyen entre los responsables de tomar decisiones, tanto internos como externos, de una organización de negocios (vea la referencia 5). Estos sistemas auditán continuamente la información contable en busca de errores y de información incompleta o inverosímil. Por ejemplo, cuando los clientes de la empresa de remodelaciones Saxon hacen un pedido en línea, el sistema de información contable revisa los formatos de pedido en busca de posibles errores. Se marcan todas las facturas cuestionables y se les incluye en el *reporte diario de excepciones*. Los datos recabados últimamente por la empresa muestran que la posibilidad de que un formato resulte marcado es de 0.10. En Saxon quieren determinar la posibilidad de encontrar cierto número de formatos marcados en una muestra de tamaño específico. Por ejemplo, en una muestra compuesta por cuatro formatos, ¿cuál sería la posibilidad de que ninguno de ellos resulte marcado? ¿Y de que uno lo sea?

¿Cómo puede la empresa de remodelaciones Saxon encontrar la solución a este problema de probabilidad? Una táctica consiste en emplear un modelo, o representación a pequeña escala, que se aproxime al proceso. Mediante el uso de esa aproximación, los directivos de Saxon podrán obtener inferencias sobre el procesamiento real de los pedidos. A pesar de que para algunos analistas la elaboración de modelos es una tarea complicada, en este caso los directivos de Saxon tienen a su disposición las *distribuciones de probabilidad*, que son modelos matemáticos apropiados para resolver el problema de probabilidad que enfrentan. La lectura de este capítulo le ayudará a aprender las características de una distribución de probabilidad y cómo aplicar específicamente las distribuciones binomial y de Poisson a los problemas financieros.

## 5.1 DISTRIBUCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA

En la sección 1.5, *variable numérica* se definió como una variable que produce respuestas numéricas tales como la cantidad de revistas a las que se suscribe o su estatura en pulgadas. Las variables numéricas se clasifican en *discretas* y *continuas*. Las variables numéricas continuas producen resultados a partir de procesos de medición; por ejemplo, su estatura. Las variables numéricas discretas producen resultados a partir de un proceso de conteo, como el número de revistas a las que se suscribe. Este capítulo trata de las distribuciones de probabilidad que representan variables numéricas discretas.

**La distribución de probabilidad para una variable aleatoria discreta** es una lista mutuamente excluyente de todos los posibles resultados numéricos para una variable aleatoria tal que una probabilidad de ocurrencia específica se asocia con cada resultado.

Por ejemplo, la tabla 5.1 ofrece la distribución de la cantidad de créditos aprobados por semana en la oficina de una sucursal bancaria local. La lista de la tabla 5.1 es colectivamente exhaustiva porque se han incluido todos los posibles resultados. Entonces, las probabilidades deben sumar 1. En la figura 5.1 aparece la representación gráfica de la tabla 5.1.

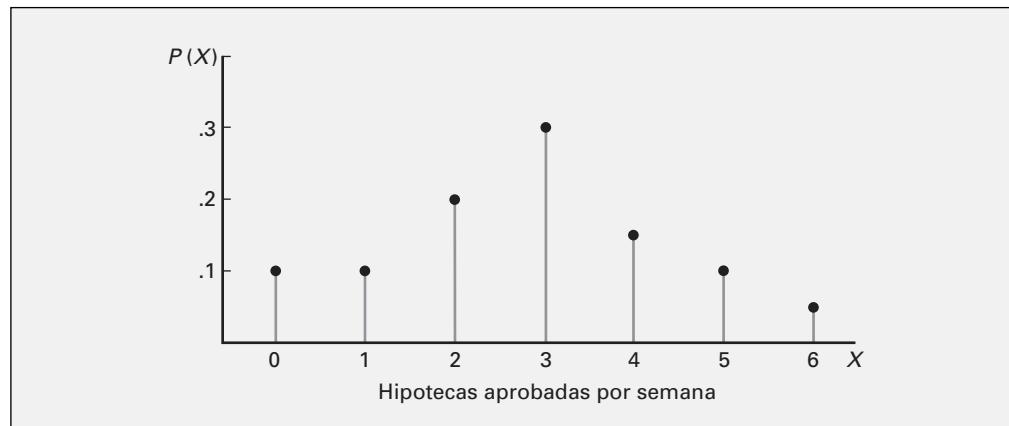
**TABLA 5.1**

Distribución de probabilidad del número de hipotecas aprobadas por semana.

Hipotecas aprobadas por semana	Probabilidad
0	0.10
1	0.10
2	0.20
3	0.30
4	0.15
5	0.10
6	0.05

**FIGURA 5.1**

Distribución de probabilidad del número de hipotecas aprobadas por semana.



### Valor esperado de una variable aleatoria discreta

La media  $\mu$  de una distribución de probabilidad es el valor esperado de su variable aleatoria. Para calcular el valor esperado, se multiplica cada resultado posible  $X$  por su probabilidad correspondiente  $P(X)$  y luego se suman estos productos.

#### VALOR ESPERADO $\mu$ DE UNA VARIABLE ALEATORIA DISCRETA

$$\mu = E(X) = \sum_{i=1}^N X_i P(X_i) \quad (5.1)$$

donde

$X_i = i$ -ésimo resultado de la variable aleatoria discreta  $X$

$P(X_i)$  = probabilidad de ocurrencia del  $i$ -ésimo resultado de  $X$

En la tabla 5.2 se calcula el valor esperado de la distribución de probabilidad del número de hipotecas aprobadas por semana (tabla 5.1), utilizando la ecuación (5.1).

**TABLA 5.2**

Cálculo del valor esperado para el número de hipotecas aprobadas por semana.

Hipotecas aprobadas por semana ( $X_i$ )	$P(X_i)$	$X_i P(X_i)$
0	0.10	$(0)(0.10) = 0.0$
1	0.10	$(1)(0.10) = 0.1$
2	0.20	$(2)(0.20) = 0.4$
3	0.30	$(3)(0.30) = 0.9$
4	0.15	$(4)(0.15) = 0.6$
5	0.10	$(5)(0.10) = 0.5$
6	0.05	$(6)(0.05) = 0.3$
	1.00	$\mu = E(X) = 2.8$

$$\begin{aligned}
\mu = E(X) &= \sum_{i=1}^N X_i P(X_i) \\
&= (0)(0.1) + (1)(0.1) + (2)(0.2) + (3)(0.3) + (4)(0.15) + (5)(0.1) + (6)(0.05) \\
&= 0 + 0.1 + 0.4 + 0.9 + 0.6 + 0.5 + 0.3 \\
&= 2.8
\end{aligned}$$

La cifra de 2.8 obtenida para el valor esperado del número de hipotecas aprobadas no es “significativa literalmente”, porque el número real de hipotecas aprobadas durante una semana determinada debe ser un valor entero. El valor esperado representa la *media* de hipotecas aprobadas por semana.

### Varianza y desviación estándar de una variable aleatoria discreta

La varianza de una distribución de probabilidad se calcula multiplicando cada posible diferencia cuadrada  $[X_i - E(X)]^2$  por su probabilidad correspondiente  $P(X_i)$  y sumando luego los productos resultantes. La ecuación (5.2) define la **varianza de una variable aleatoria discreta**.

#### VARIANZA DE UNA VARIABLE ALEATORIA DISCRETA

$$\sigma^2 = \sum_{i=1}^N [X_i - E(X)]^2 P(X_i) \quad (5.2)$$

donde

$X_i = i$ -ésimo resultado de la variable aleatoria discreta  $X$

$P(X_i) =$  probabilidad de ocurrencia del  $i$ -ésimo resultado de  $X$

La ecuación (5.3) define la **desviación estándar de una variable aleatoria discreta**.

#### DESVIACIÓN ESTÁNDAR DE UNA VARIABLE ALEATORIA DISCRETA

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^N [X_i - E(X)]^2 P(X_i)} \quad (5.3)$$

En la tabla 5.3 se calculan la varianza y la desviación estándar del número de hipotecas aprobadas por semana, mediante las ecuaciones (5.2) y (5.3).

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^N [X_i - E(X)]^2 P(X_i) \\
&= (0 - 2.8)^2(0.10) + (1 - 2.8)^2(0.10) + (2 - 2.8)^2(0.20) + (3 - 2.8)^2(0.30) \\
&\quad + (4 - 2.8)^2(0.15) + (5 - 2.8)^2(0.10) + (6 - 2.8)^2(0.05) \\
&= 0.784 + 0.324 + 0.128 + 0.012 + 0.216 + 0.484 + 0.512 \\
&= 2.46
\end{aligned}$$

y

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.46} = 1.57$$

Así, la media de hipotecas aprobadas por semana es 2.8, la varianza es 2.46 y la desviación estándar es 1.57.

**TABLA 5.3**

Cálculo de la varianza y la desviación estándar del número de hipotecas aprobadas por semana.

Hipotecas aprobadas por semana ( $X_i$ )	Probabilidad		
	$P(X_i)$	$X_i P(X_i)$	$[X_i - E(X)]^2 P(X_i)$
0	0.10	$(0)(0.10) = 0.0$	$(0 - 2.8)^2(0.10) = 0.784$
1	0.10	$(1)(0.10) = 0.1$	$(1 - 2.8)^2(0.10) = 0.324$
2	0.20	$(2)(0.20) = 0.4$	$(2 - 2.8)^2(0.20) = 0.128$
3	0.30	$(3)(0.30) = 0.9$	$(3 - 2.8)^2(0.30) = 0.012$
4	0.15	$(4)(0.15) = 0.6$	$(4 - 2.8)^2(0.15) = 0.216$
5	0.10	$(5)(0.10) = 0.5$	$(5 - 2.8)^2(0.10) = 0.484$
6	0.05	$(6)(0.05) = 0.3$	$(6 - 2.8)^2(0.05) = 0.512$
$\sigma^2 = \sum_{i=1}^N [X_i - E(X)]^2 P(X_i) = 2.46$			
$\sigma = 1.57$			

## PROBLEMAS PARA LA SECCIÓN 5.1

### Aprendizaje básico

**ASISTENCIA de PH Grade** **5.1** Dadas las siguientes distribuciones de probabilidad:

Distribución A		Distribución B	
$X$	$P(X)$	$X$	$P(X)$
0	0.50	0	0.05
1	0.20	1	0.10
2	0.15	2	0.15
3	0.10	3	0.20
4	0.05	4	0.50

- Calcule el valor esperado de cada distribución.
- Calcule la desviación estándar de cada distribución.
- Compare e identifique las diferencias entre los resultados de las distribuciones *A* y *B*.

**ASISTENCIA de PH Grade** **5.2** Dadas las siguientes distribuciones de probabilidad:

Distribución C		Distribución D	
$X$	$P(X)$	$X$	$P(X)$
0	0.20	0	0.10
1	0.20	1	0.20
2	0.20	2	0.40
3	0.20	3	0.20
4	0.20	4	0.10

- Calcule el valor esperado de cada distribución.
- Calcule la desviación estándar de cada distribución.
- Compare e identifique las diferencias entre los resultados de las distribuciones *C* y *D*.

### Aplicación de conceptos

**ASISTENCIA de PH Grade** **5.3** Utilizando los registros de la empresa correspondientes a los últimos 500 días hábiles, el gerente

de Konig Motors, concesionario automotriz suburbano, sintetizó el número de automóviles vendidos al día en la siguiente tabla:

Número de automóviles vendidos al día	Frecuencia de ocurrencia
0	40
1	100
2	142
3	66
4	36
5	30
6	26
7	20
8	16
9	14
10	8
11	2
Total	500

- Construya la distribución de probabilidad para el número de automóviles vendidos al día.
  - Calcule la media o el número esperado de automóviles vendidos al día.
  - Calcule la desviación estándar.
- 5.4** En la siguiente tabla se encuentra la distribución de probabilidad para el número diario de accidentes de tráfico ocurridos en una ciudad pequeña.

Número diario de accidentes ( $X$ )	$P(X)$
0	0.10
1	0.20
2	0.45
3	0.15
4	0.05
5	0.05

- Calcule la media o el número esperado de accidentes ocurridos al día.
- Calcule la desviación estándar.



- 5.5** El gerente de un sistema de redes de computadoras desarrolló la siguiente distribución de probabilidad para el número de interrupciones al día:

Interrupciones (X)	P(X)
0	0.32
1	0.35
2	0.18
3	0.08
4	0.04
5	0.02
6	0.01

- Calcule la media o el número esperado de interrupciones por día.
- Calcule la desviación estándar.

- 5.6** En el juego de feria *Más o menos de siete*, se tira un par de dados sin cargar, y la suma resultante determina si el jugador

gana o pierde su apuesta. Por ejemplo, el jugador puede apostar \$1.00 a que la suma será menor a 7, es decir, 2, 3, 4, 5 o 6. En este caso, el jugador perderá \$1.00 si el resultado es igual o mayor que 7, o lo ganará si el resultado es menor que 7. Del mismo modo, puede apostar \$1.00 a que la suma será mayor que 7, es decir, 8, 9, 10, 11 o 12. Aquí, gana \$1.00 si el resultado es mayor que 7, pero lo pierde si el resultado es 7 o menos. Una tercera opción del juego consiste en apostar \$1.00 a que el resultado es 7. En esta apuesta, el jugador ganará \$4.00 si el resultado del tiro es 7 y perderá \$1.00 si es cualquier otro.

- Elabore la distribución de probabilidad que representa los distintos resultados posibles para una apuesta de \$1.00 a que el tiro será menor que 7.
- Elabore la distribución de probabilidad que representa los distintos resultados posibles para una apuesta de \$1.00 a que el tiro será mayor que 7.
- Elabore la distribución de probabilidad que representa a los distintos resultados posibles para una apuesta de \$1.00 a que el tiro será de 7.
- Demuestre que las ganancias (o pérdidas) a largo plazo esperadas del jugador son iguales, independientemente del método de juego que utilice.

## 5.2 DISTRIBUCIÓN BINOMIAL

En los siguientes dos apartados se utilizan modelos matemáticos para resolver problemas financieros.

Un **modelo matemático** es una expresión matemática que se utiliza para representar una variable de interés.

Cuando se dispone de una expresión matemática, es factible calcular la probabilidad de ocurrencia exacta correspondiente a cualquier resultado específico para la variable aleatoria.

La **distribución de probabilidad binomial** es uno de los modelos matemáticos más útiles. La distribución binomial se utiliza cuando la variable aleatoria discreta de interés es el número de éxitos en una muestra compuesta por  $n$  observaciones. La distribución binomial tiene cuatro propiedades fundamentales:

- La muestra se compone de un número fijo de observaciones,  $n$ .
- Cada observación se clasifica en una de dos categorías mutuamente excluyentes y colectivamente exhaustivas, normalmente denominadas *éxito* y *fracaso*.
- La probabilidad de que una observación se clasifique como éxito,  $p$ , es constante de una observación a otra. De la misma forma, la probabilidad de que una observación se clasifique como fracaso,  $1 - p$ , es constante en todas las observaciones.
- El resultado (es decir, el éxito o el fracaso) de cualquier observación es independiente del resultado de cualquier otra observación. Para garantizar la independencia, las observaciones se deben seleccionar de manera aleatoria, ya sea de una *población infinita sin reemplazo* o de una *población finita con reemplazo*.

Volviendo al escenario “Uso de la estadística” expuesto en la página 154 referente al sistema de información contable, supongamos que *éxito* se define como un formato de pedido marcado y *fracaso* es cualquier otro resultado. Usted se interesa por el número de formatos marcados dentro de una muestra de pedidos determinada.

¿Qué resultados se pueden presentar? Si la muestra se compone de cuatro pedidos, puede haber ninguna, una, dos, tres o cuatro formatos marcados. La variable aleatoria binomial, que es el número de formatos de pedido marcados, no puede asumir ningún otro valor porque el número de formatos de pedido marcados no podrá ser superior al tamaño  $n$  de la muestra ni podrá ser menor que cero. Por lo tanto, la variable aleatoria binomial tiene un rango de 0 a  $n$ .

Suponga que en una muestra de cuatro pedidos, usted observa el siguiente resultado:

Primer pedido	Segundo pedido	Tercer pedido	Cuarto pedido
Marcado	Marcado	Sin marcar	Marcado

¿Cuál es la probabilidad de tener tres éxitos (formatos de pedido marcados) con esta secuencia en particular, en una muestra compuesta por cuatro pedidos? Puesto que la probabilidad histórica de un pedido marcado es de 0.10, la probabilidad de que cada pedido se presente con dicha secuencia es

Primer pedido	Segundo pedido	Tercer pedido	Cuarto pedido
$p = 0.10$	$p = 0.10$	$1 - p = 0.90$	$p = 0.10$

Cada uno de los resultados es independiente de los demás, ya que los formatos de pedido se seleccionaron de una población extremadamente grande o prácticamente infinita y sin reemplazarlos. Por tanto, la probabilidad de tener esta secuencia en particular es

$$\begin{aligned}
 pp(1 - p)p &= p^3(1 - p)^1 \\
 &= (0.10)(0.10)(0.10)(0.90) \\
 &= (0.10)^3(0.90)^1 \\
 &= 0.0009
 \end{aligned}$$

Este resultado sólo señala la probabilidad de que aparezcan tres formatos de pedido marcados (éxitos) extraídos de una muestra de cuatro con una *secuencia específica*. Para encontrar cuántas maneras hay de seleccionar  $X$  objetos de entre  $n$  objetos, *independientemente de la secuencia*, se utiliza la **regla de las combinaciones** dada en la ecuación (5.4).

### COMBINACIONES

El número de combinaciones para seleccionar  $X$  objetos de entre  $n$  objetos está dado por:

$${}_n C_X = \frac{n!}{X!(n - X)!} \quad (5.4)$$

donde  $n! = (n)(n - 1) \dots (1)$  se llama  $n$  factorial. Por definición,  $0! = 1$ .

Con  $n = 4$  y  $X = 3$ , hay

$${}_n C_X = \frac{n!}{X!(n - X)!} = \frac{4!}{3!(4 - 3)!} = \frac{4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(1)} = 4$$

de tales secuencias. Las cuatro secuencias posibles son:

**Secuencia 1** = marcada, marcada, marcada, sin marcar con una probabilidad

$$ppp(1 - p) = p^3(1 - p)^1 = 0.0009$$

**Secuencia 2** = marcada, marcada, sin marcar, marcada con una probabilidad

$$pp(1 - p)p = p^3(1 - p)^1 = 0.0009$$

**Secuencia 3** = marcada, sin marcar, marcada, marcada con una probabilidad

$$p(1 - p)pp = p^3(1 - p)^1 = 0.0009$$

**Secuencia 4** = sin marcar, marcada, marcada, marcada con una probabilidad

$$(1 - p)ppp = p^3(1 - p)^1 = 0.0009$$

Por tanto, la probabilidad de que haya tres formatos de pedido marcados es igual a

$$\begin{aligned}
 (\text{número de secuencias posibles}) \times (\text{probabilidad de una secuencia en particular}) &= (4) \times \\
 (0.0009) &= 0.0036
 \end{aligned}$$

Usted puede hacer una deducción intuitiva similar para los demás resultados posibles de la variable aleatoria: cero, uno, dos y cuatro formatos de pedido marcados. Sin embargo, a medida que  $n$ , el tamaño de la muestra, es mayor, los cálculos implicados para este método intuitivo requieren de más tiempo. Existe un modelo matemático que brinda una fórmula general para calcular cualquier probabilidad binomial. La ecuación (5.5) es el modelo matemático que se utiliza para representar la distribución de probabilidad binomial y calcular el número de éxitos ( $X$ ), dados los valores  $n$  y  $p$ .

### DISTRIBUCIÓN DE PROBABILIDAD BINOMIAL

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \quad (5.5)$$

donde

$P(X)$  = probabilidad de  $X$  éxitos, dadas  $n$  y  $p$

$n$  = número de observaciones

$p$  = probabilidad de éxitos

$1 - p$  = probabilidad de fracasos

$X$  = número de éxitos en la muestra ( $X = 0, 1, 2, \dots, n$ )

La ecuación (5.5) replantea lo que usted dedujo de manera intuitiva. La variable aleatoria binomial  $X$  puede tener cualquier valor entero  $X$  desde 0 hasta  $n$ . En la ecuación (5.5) el producto

$$p^X (1-p)^{n-X}$$

indica la probabilidad de obtener exactamente  $X$  éxitos de entre  $n$  observaciones en una *secuencia específica*. El término:

$$\frac{n!}{X!(n-X)!}$$

indica *cuántas combinaciones* de esos  $X$  éxitos de entre  $n$  observaciones son posibles. Así, dado el número de observaciones  $n$  y la probabilidad de éxito  $p$ , la probabilidad de obtener  $X$  éxitos es:

$P(X)$  = (número de secuencias posibles)  $\times$  (probabilidad de una secuencia específica)

$$= \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

En el ejemplo 5.1 se ilustra el uso de la ecuación (5.5).

### EJEMPLO 5.1

#### DETERMINAR $P(X = 3)$ , DADAS $n = 4$ Y $p = 0.1$

Si la posibilidad de que un formato de pedido sea marcado es de 0.1, ¿qué probabilidad existe de que haya tres formatos marcados en una muestra de cuatro?

**SOLUCIÓN** Al utilizar la ecuación (5.5), se sabe que la probabilidad de obtener tres formatos marcados de una muestra compuesta por cuatro es

$$\begin{aligned} P(X = 3) &= \frac{4!}{3!(4-3)!} (0.1)^3 (1-0.1)^{4-3} \\ &= \frac{4!}{3!(4-3)!} (0.1)^3 (0.9)^1 \\ &= 4(0.1)(0.1)(0.1)(0.9) = 0.0036 \end{aligned}$$

En los ejemplos 5.2 y 5.3 se muestran los cálculos para otros valores de  $X$ .

**EJEMPLO 5.2**DETERMINAR  $P(X \geq 3)$ , DADAS  $n = 4$  Y  $p = 0.1$ 

Si la posibilidad de que un formato de pedido sea marcado es de 0.1, ¿qué probabilidad existe de que haya tres o más formatos marcados (es decir, por lo menos tres) en una muestra de cuatro?

**SOLUCIÓN** En el ejemplo 5.1 usted encontró que la probabilidad de obtener *exactamente* tres formatos marcados a partir de una muestra de cuatro es 0.0036. Para calcular la probabilidad de obtener *por lo menos* tres formatos marcados, es necesario sumar la probabilidad de los tres formatos marcados y la probabilidad de cuatro formatos marcados. La probabilidad de obtener cuatro formatos marcados es:

$$\begin{aligned} P(X = 4) &= \frac{4!}{4!(4-4)!}(0.1)^4(1-0.1)^{4-4} \\ &= \frac{4!}{4!(0)!}(0.1)^4(0.9)^0 \\ &= 1(0.1)(0.1)(0.1)(0.1) = 0.0001 \end{aligned}$$

Así, la probabilidad de obtener al menos tres formatos de pedido marcados es:

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) \\ &= 0.0036 + 0.0001 \\ &= 0.0037 \end{aligned}$$

Existe un 0.37% de posibilidad de tener al menos tres formatos de pedido marcados en una muestra compuesta por cuatro formatos.

**EJEMPLO 5.3**DETERMINAR  $P(X < 3)$ , DADAS  $n = 4$  Y  $p = 0.1$ 

Si la posibilidad de que un formato de pedido sea marcado es de 0.1, ¿qué probabilidad existe de que haya menos de tres formatos marcados en una muestra compuesta por cuatro?

**SOLUCIÓN** La probabilidad de obtener menos de tres formatos de pedido marcados es:

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

Se utiliza la ecuación (5.5) de la página 160 para calcular cada una de las probabilidades:

$$\begin{aligned} P(X = 0) &= \frac{4!}{0!(4-0)!}(0.1)^0(1-0.1)^{4-0} = 0.6561 \\ P(X = 1) &= \frac{4!}{1!(4-1)!}(0.1)^1(1-0.1)^{4-1} = 0.2916 \\ P(X = 2) &= \frac{4!}{2!(4-2)!}(0.1)^2(1-0.1)^{4-2} = 0.0486 \end{aligned}$$

Por tanto,  $P(X < 3) = 0.6561 + 0.2916 + 0.0486 = 0.9963$ .

$P(X < 3)$  también se puede calcular a partir de su complemento  $P(X \geq 3)$ , como sigue:

$$\begin{aligned} P(X < 3) &= 1 - P(X \geq 3) \\ &= 1 - 0.0037 = 0.9963 \end{aligned}$$

Tal vez los cálculos realizados en el ejemplo 5.3 resulten tediosos, sobre todo conforme  $n$  aumenta. Para evitar la monotonía que implica efectuar los cálculos, es factible encontrar muchas probabilidades binomiales directamente en la tabla E.6, que se reproduce parcialmente en la tabla 5.4. La tabla

E.6 proporciona las probabilidades binomiales correspondientes a  $X = 0, 1, 2, \dots, n$ , para una selección de varias combinaciones de  $n$  y  $p$ . Por ejemplo, para encontrar la probabilidad de obtener exactamente dos éxitos en una muestra de cuatro, cuando la probabilidad de éxito es 0.1, vea primero  $n = 4$  y luego busque en la fila  $X = 2$  y en la columna  $p = 0.10$ . El resultado es 0.0486.

**TABLA 5.4**

Cómo encontrar la probabilidad binomial para  $n = 4$ ,  $X = 2$  y  $p = 0.1$ .

<b><i>n</i></b>	<b><i>X</i></b>	<b>0.01</b>	<b>0.02</b>	<b><i>p</i></b>	
4	0	0.9606	0.9224	....	0.6561
	1	0.0388	0.0753	....	0.2916
	2	0.0006	0.0023	....	0.0486
	3	0.0000	0.0000	....	0.0036
	4	0.0000	0.0000	....	0.0001

Fuente: Tabla E.6.

También es posible calcular las probabilidades binomiales que aparecen en la tabla E.6 utilizando Excel o Minitab. En la figura 5.2 aparece una hoja de trabajo de Excel para calcular probabilidades binomiales, y en la figura 5.3 se ilustra el resultado en Minitab.

**FIGURA 5.2**

Hoja de trabajo de Excel para calcular probabilidades binomiales.

A	B
1 Tagged Orders	
2	
3 Data	
4 Sample size	4
5 Probability of success	0.1
6	
7 Statistics	
8 Mean	0.4
9 Variance	0.36
10 Standard deviation	0.6
11	
12 Binomial Probabilities Table	
13	X P(X)
14	0 0.6561
15	1 0.2916
16	2 0.0486
17	3 0.0036
18	4 0.0001

=B4 \* B5  
 =B8 \* (1 - B5)  
 =SQRT(B9)

=BINOMDIST(A14, \$B\$4, \$B\$5, FALSE)  
 =BINOMDIST(A15, \$B\$4, \$B\$5, FALSE)  
 =BINOMDIST(A16, \$B\$4, \$B\$5, FALSE)  
 =BINOMDIST(A17, \$B\$4, \$B\$5, FALSE)  
 =BINOMDIST(A18, \$B\$4, \$B\$5, FALSE)

**FIGURA 5.3**

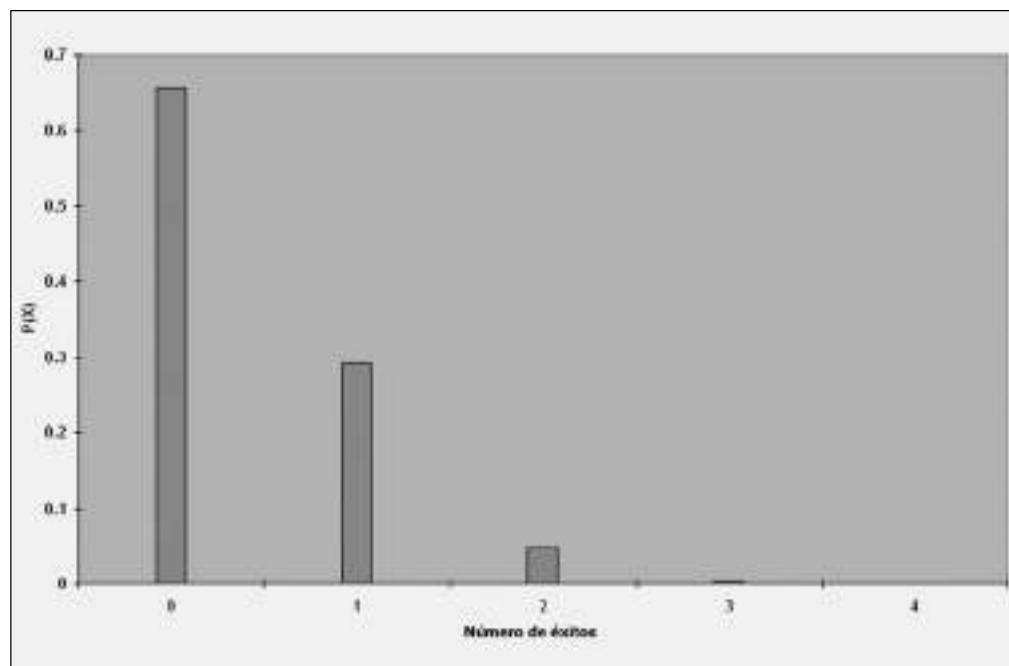
Cálculo de la distribución binomial para  $n = 4$  y  $p = 0.1$  en Minitab.

Binomial with $n = 4$ and $p = 0.1$	
x	$P(X = x)$
0	0.6561
1	0.2916
2	0.0486
3	0.0036
4	0.0001

La forma de probabilidad binomial depende de los valores de  $n$  y  $p$ . Siempre que  $p = 0.5$ , la distribución binomial es simétrica, independientemente de lo grande o pequeño del valor de  $n$ . Cuando  $p \neq 0.5$ , la distribución es asimétrica. Cuanto más se acerca  $p$  a 0.5 y el número de observaciones  $n$  es mayor, menos asimétrica se vuelve la distribución. Por ejemplo, la distribución del número de formatos marcados es muy asimétrica a la derecha, porque  $p = 0.1$  y  $n = 4$  (vea la figura 5.4).

**FIGURA 5.4**

Gráfica de barras de Excel de la distribución de probabilidad binomial con  $n = 4$  y  $p = 0.1$ .



La media de la distribución binomial es igual al producto de  $n$  por  $p$ . En lugar de calcular la media de la distribución de probabilidad utilizando la ecuación (5.1) de la página 155, utilice la ecuación (5.6) para calcular la media de las variables que siguen a la distribución binomial.

#### MEDIA DE LA DISTRIBUCIÓN BINOMIAL

La media  $\mu$  de la distribución binomial es igual a la multiplicación del tamaño  $n$  de la muestra por la probabilidad de éxito  $p$ .

$$\mu = E(X) = np \quad (5.6)$$

En promedio, y a la larga, teóricamente usted esperaría  $\mu = E(X) = np = (4)(0.1) = 0.4$  formatos de pedido marcados en una muestra compuesta por cuatro formatos.

La desviación estándar de la distribución binomial se calcula mediante la ecuación (5.7).

#### DESVIACIÓN ESTÁNDAR DE LA DISTRIBUCIÓN BINOMIAL

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{np(1-p)} \quad (5.7)$$

La desviación estándar del número de formatos marcados es:

$$\sigma = \sqrt{4(0.1)(0.9)} = 0.6$$

Este resultado es el mismo que calcularía si utilizara la ecuación (5.3) de la página 156.

#### EJEMPLO 5.4

#### CÁLCULO DE PROBABILIDADES BINOMIALES

La exactitud al tomar los pedidos en la ventanilla de servicio a los automovilistas es una característica muy importante de las cadenas de comida rápida. Todos los meses, *QSR Magazine* [www.qsrmagazine.com](http://www.qsrmagazine.com) publica los resultados de sus encuestas. La exactitud se evalúa como el porcentaje de pedidos compuestos por un artículo principal, uno secundario y una bebida (pero solicitando

que se elimine algún ingrediente normal como los pepinillos) que se sirven correctamente. Hace poco, el porcentaje de pedidos de este tipo servidos correctamente en Burger King fue del 88%. Suponga que usted y dos de sus amigos van en su automóvil a la ventanilla de servicio de Burger King y cada uno hace un pedido como el descrito antes. ¿Cuál es la probabilidad de que los tres pedidos se sirvan con exactitud? ¿Ninguno de los tres? ¿Al menos dos de los tres? ¿Cuáles son el promedio y la desviación estándar del número de pedidos servidos con exactitud?

**SOLUCIÓN** Puesto que se trata de tres pedidos y la probabilidad de tener un pedido servido con exactitud es del 88%,  $n = 3$  y  $p = 0.88$ . Mediante las ecuaciones (5.5), (5.6) y (5.7)

$$\begin{aligned} P(X = 3) &= \frac{3!}{3!(3-3)!} (0.88)^3 (1-0.88)^{3-3} \\ &= \frac{3!}{3!(3-3)!} (0.88)^3 (0.12)^0 \\ &= 1(0.88)(0.88)(0.88)(1) = 0.6815 \end{aligned}$$

$$\begin{aligned} P(X = 0) &= \frac{3!}{0!(3-0)!} (0.88)^0 (1-0.88)^{3-0} \\ &= \frac{3!}{0!(3-0)!} (0.88)^0 (0.12)^3 \\ &= 1(1)(0.12)(0.12)(0.12) = 0.0017 \end{aligned}$$

$$\begin{aligned} P(X = 2) &= \frac{3!}{2!(3-2)!} (0.88)^2 (1-0.88)^{3-2} \\ &= \frac{3!}{2!(3-2)!} (0.88)^2 (0.12)^1 \\ &= 3(0.88)(0.88)(0.12) = 0.2788 \end{aligned}$$

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) \\ &= 0.2788 + 0.6815 \\ &= 0.9603 \end{aligned}$$

$$\mu = E(X) = np = 3(0.88) = 2.64$$

$$\begin{aligned} \sigma &= \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{np(1-p)} \\ &= \sqrt{3(0.88)(0.12)} \\ &= \sqrt{0.3168} = 0.563 \end{aligned}$$

La probabilidad de que los tres pedidos se sirvan con exactitud es de 0.6815 o del 68.15%. La probabilidad de que ninguno de los pedidos se sirva con exactitud es de 0.0017 o del 0.17%. La probabilidad de que por lo menos dos pedidos se sirvan con exactitud es de 0.9603 o del 96.03%. La media de pedidos servidos con exactitud en una muestra de tres pedidos es 2.64 y la desviación estándar es 0.563.

En este apartado se explicó la distribución binomial, la cual desempeña un papel aún más importante cuando se utiliza en problemas de inferencia estadística que implican la estimación o prueba de hipótesis sobre proporciones (como analizaremos en los capítulos 8 y 9).

## PROBLEMAS PARA LA SECCIÓN 5.2

Puede resolver manualmente los problemas 5.7 a 5.14 o usando Excel o Minitab. Le recomendamos usar Excel o Minitab para resolver los problemas 5.15 a 5.17.

AUTO  
Examen

### Aprendizaje básico

**5.7** Si  $n = 5$  y  $p = 0.40$ , encuentre cuál es la probabilidad de que

- $X = 4$
- $X \leq 3$
- $X < 2$
- $X > 1$

ASISTENCIA  
de PH Grade

**5.8** Determine lo siguiente:

- Para  $n = 4$  y  $p = 0.12$ , ¿cuánto es  $P(X = 0)$ ?
- Para  $n = 10$  y  $p = 0.40$ , ¿cuánto es  $P(X = 9)$ ?
- Para  $n = 10$  y  $p = 0.50$ , ¿cuánto es  $P(X = 8)$ ?
- Para  $n = 6$  y  $p = 0.83$ , ¿cuánto es  $P(X = 5)$ ?

### Aplicación de conceptos

**5.9** Se supone que el aumento o la disminución del precio de una acción durante el transcurso de un día hábil es un evento aleatorio igualmente posible. ¿Cuál es la probabilidad de que una acción muestre un aumento en su precio al cierre durante cinco días consecutivos?

**5.10** El 60% de los estadounidenseslean su contrato de trabajo, incluyendo las letras pequeñas (“Snapshots,” **usatoday.com**, 20 de enero, 2004). Suponga que el número de empleados quelean cada una de las palabras de su contrato se puede modelar utilizando la distribución binomial. Considerando un grupo de cinco empleados, encuentre cuál es la probabilidad de que:

- Los cinco lean cada una de las palabras de su contrato.
- Al menos tres lean cada una de las palabras de su contrato.
- Menos de dos lean cada una de las palabras de su contrato.
- ¿Cuáles serían sus respuestas para los incisos a) a c), si la probabilidad de que un empleado lea cada una de las palabras de su contrato es de 0.80?

ASISTENCIA  
de PH Grade

**5.11** Una estudiante presenta un examen de opción múltiple, en el que cada pregunta tiene cuatro opciones. Suponga que ella no conoce la respuesta correcta de ninguna de las preguntas, y que decidió utilizar una estrategia en la que colocará cuatro pelotas (marcadas como A, B, C y D) dentro de una caja. Ahora, para responder cada pregunta, selecciona una pelota de manera aleatoria y la devuelve a la caja. La letra de la pelota determinará su respuesta a la pregunta. El examen se compone de cinco preguntas de opción múltiple. ¿Cuál es la probabilidad de que ella obtenga

- cinco respuestas correctas?
- al menos cuatro respuestas correctas?
- ninguna respuesta correcta?
- no más de dos respuestas correctas?

**5.12** En el ejemplo 5.4 de la página 163, usted y dos amigos decidieron ir a Burger King. En lugar de eso, suponga que acuden a McDonald’s, donde durante el mes pasado sirvieron el 90% de los pedidos con exactitud. ¿Cuál es la probabilidad de que

- los tres pedidos se sirvan con exactitud?
- ninguno de los tres pedidos se sirva con exactitud?
- al menos dos de los tres pedidos se sirva con exactitud?
- ¿Cuáles son la media y la desviación estándar del número de pedidos servidos con exactitud?

**5.13** Durante varios años se ha reducido el porcentaje de comisión que las líneas aéreas comerciales pagan a los agentes de viajes. Muchas agencias, en busca de mejorar sus ingresos, cobran ahora a sus clientes una cuota por boleto, generalmente de entre 10 y 25 dólares. De acuerdo con la sociedad estadounidense de agentes boleteros, cerca del 90% de los agentes de viajes cobra cuotas a sus clientes cuando éstos adquieren boletos de avión (Kortney Stringer, “American Air Fees for Travel Agents to Be Cut Again”, *The Wall Street Journal*, 20 de agosto, 2001, B2).

- La cifra del 90% citada por la sociedad estadounidense de agentes boleteros, ¿quedaría mejor clasificada como probabilidad clásica *a priori*, probabilidad clásica empírica o probabilidad subjetiva?
- Usted selecciona una muestra aleatoria de 10 agencias de viajes. Suponga que el número de 10 agencias de viajes que cobran una cuota por boleto se distribuye como variable aleatoria binomial. ¿Cuáles son la media y la desviación estándar de esta distribución?
- ¿Qué suposiciones es necesario hacer en el inciso b)?

**5.14** Consulte el problema 5.13 y calcule la probabilidad de que de 10 agencias de viajes:

- ninguna cobre cuota por boleto.
- exactamente una cobre cuota por boleto.
- dos o menos cobren cuota por boleto.
- tres o más cobren cuota por boleto.

ASISTENCIA  
de PH Grade

**5.15** Cuando un cliente hace un pedido a la Papelería en Línea de Rudy, un sistema contable computarizado (AIS, por sus siglas en inglés) verifica automáticamente si el cliente ha excedido o no su límite de crédito. Los registros señalan que la probabilidad de que los clientes exceden su límite de crédito es de 0.05. Suponga que durante un día determinado, 20 clientes hicieron un pedido. Suponga también que el número de clientes que según el sistema AIS excedieron su límite de crédito está distribuido como variable aleatoria binomial.

- ¿Cuáles son la media y la desviación estándar del número de clientes que excedieron su límite de crédito?
- ¿Cuál es la probabilidad de que ningún cliente exceda su límite de crédito?
- ¿Cuál es la probabilidad de que sólo un cliente exceda su límite de crédito?
- ¿Cuál es la probabilidad de que dos o más clientes excedan su límite de crédito?

**5.16** Cada otoño las televisoras presentan nuevos programas. Con la idea de atraer el interés de los espectadores, durante el verano se transmiten anuncios como parte de una campaña publicitaria previa a su aparición en otoño. Después, las televisoras realizan encuestas para ver qué porcentaje de los espectadores *está al corriente* de los estrenos. De acuerdo con datos procedentes de las televisoras, durante el otoño de 2001, el 68% de los espectadores de entre 18 y 49 años supieron de la nueva serie *Criminal Intent*, mientras que sólo el 24% de ellos escucharon de *Inside Schwartz* (Joe Flint, “Viewers Awareness of New Shows Rises”, *The Wall Street Journal*, 20 de agosto, 2001, B7).

a. Las cifras del 68% y del 24% citadas por las televisoras, ¿quedarían mejor clasificadas como probabilidades clásicas *a priori*, probabilidades clásicas empíricas o probabilidades subjetivas?

Suponga que selecciona una muestra de 20 espectadores con edades de 18 a 49 años. ¿Cuál es la probabilidad de que:

- menos de cinco espectadores vean *Criminal Intent*?
- 10 o más espectadores vean *Criminal Intent*?
- los 20 vean *Criminal Intent*?

**5.17** De acuerdo con los datos del problema 5.16, se obtiene otra muestra de 20 espectadores y se considera al nuevo programa *Inside Schwartz*. ¿Cuál es la probabilidad de que:

- menos de cinco espectadores vean *Inside Schwartz*?
- 10 o más espectadores vean *Inside Schwartz*?
- los 20 vean *Inside Schwartz*?
- Compare los resultados de los incisos a) a c) con los de *Criminal Intent* del problema 5.16, incisos b) a d).

### 5.3 DISTRIBUCIÓN DE POISSON

Muchos estudios se basan en el conteo de las veces que se presenta un evento dentro de un *área de oportunidad* dada. El **área de oportunidad** es una unidad continua o intervalo de tiempo, volumen, o área en donde se puede presentar más de un evento. Algunos ejemplos serían los defectos en la superficie de un refrigerador, el número de fallas de la red en un día, o el número de pulgas que tiene un perro. Cuando se tiene un área de oportunidad como éstas, se utiliza la **distribución de Poisson** para calcular las probabilidades si:

- Le interesa contar las veces que se presenta un evento en particular dentro de un área de oportunidad determinada. El área de oportunidad se define por tiempo, extensión, área, etcétera.
- La probabilidad de que un evento se presente en un área de oportunidad dada es igual para todas las áreas de oportunidad.
- El número de eventos que ocurren en un área de oportunidad es independiente del número de eventos que se presentan en cualquier otra área de oportunidad.
- La probabilidad de que dos o más eventos se presenten en un área de oportunidad tiende a cero conforme esa área se vuelve menor.

Considere el número de clientes que llegan a un banco ubicado en la zona central de negocios de una gran ciudad, durante la hora del almuerzo. A usted le interesa conocer el número de clientes que llegan cada minuto. ¿Esta situación concuerda con las cuatro propiedades de la distribución de Poisson mencionadas anteriormente? Primero, el *evento* de interés es un cliente y el *área de oportunidad* está definida como un intervalo de 1 minuto. ¿Llegarán cero clientes, un cliente, dos clientes, etcétera? Segundo, es razonable suponer que la probabilidad de que llegue un cliente durante un intervalo específico de 1 minuto es igual a la probabilidad correspondiente a todos los demás intervalos de 1 minuto. Tercero, la llegada de un cliente durante cualquier intervalo de 1 minuto no influye en (es decir, es estadísticamente independiente de) la llegada de cualquier otro cliente durante cualquier otro intervalo de 1 minuto. Por último, la probabilidad de que lleguen dos o más clientes durante un periodo de tiempo dado tiende a cero a medida que dicho intervalo de tiempo es menor. Por ejemplo, la probabilidad de que dos clientes lleguen durante un intervalo de tiempo con una amplitud de 1/100 de segundo es virtualmente cero. Así, es factible utilizar la distribución de Poisson para determinar las probabilidades correspondientes al número de clientes que llegan al banco durante un intervalo de tiempo de 1 minuto mientras es la hora del almuerzo.

La distribución de Poisson tiene un parámetro, llamado  $\lambda$  (la letra griega *lambda* minúscula), que es la media o el número esperado de eventos por unidad. La varianza de la distribución de Poisson también es igual a  $\lambda$ , y su desviación estándar es igual a  $\sqrt{\lambda}$ . El número de eventos  $X$  de la variable aleatoria de Poisson fluctúa desde 0 hasta infinito.

En la ecuación (5.8) se presenta la expresión matemática de la distribución de Poisson para el cálculo de la probabilidad de  $X$  eventos, dado que se esperan  $\lambda$  eventos.

### DISTRIBUCIÓN DE LA PROBABILIDAD DE POISSON

$$P(X) = \frac{e^{-\lambda} \lambda^X}{X!} \quad (5.8)$$

donde  $P(X)$  = la probabilidad de  $X$  eventos en un área de oportunidad

$\lambda$  = número de eventos esperado

$e$  = constante matemática aproximadamente igual a 2.71828

$X$  = número de eventos

Para demostrar la distribución de Poisson, suponga que la media de clientes que llega al banco por minuto durante la hora que va del mediodía a la 1 PM es igual a 3.0. ¿Cuál es la probabilidad de que lleguen exactamente dos clientes durante un minuto dado? ¿Y cuál es la probabilidad de que lleguen más de dos clientes durante un minuto dado?

Al utilizar la ecuación (5.8) y  $\lambda = 3$ , se sabe que la probabilidad de que lleguen exactamente dos clientes durante el transcurso de un minuto dado es

$$P(X = 2) = \frac{e^{-3.0}(3.0)^2}{2!} = \frac{9}{(2.71828)^3(2)} = 0.2240$$

Para determinar la probabilidad de que lleguen más de dos clientes durante un minuto dado

$$P(X > 2) = P(X = 3) + P(X = 4) + \cdots + P(X = \infty)$$

Puesto que la suma de todas las probabilidades de una distribución de probabilidad debe ser igual a 1, los términos del lado derecho de la ecuación  $P(X > 2)$  también representan al complemento de la probabilidad de que  $X$  sea menor o igual que 2 [es decir,  $1 - P(X \leq 2)$ ]. Así,

$$P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

Ahora, con la ecuación (5.8),

$$\begin{aligned} P(X > 2) &= 1 - \left[ \frac{e^{-3.0}(3.0)^0}{0!} + \frac{e^{-3.0}(3.0)^1}{1!} + \frac{e^{-3.0}(3.0)^2}{2!} \right] \\ &= 1 - [0.0498 + 0.1494 + 0.2240] \\ &= 1 - 0.4232 = 0.5768 \end{aligned}$$

De esta forma, existe una posibilidad del 57.68% de que lleguen más de dos clientes en el mismo minuto.

Para evitar la monotonía que implica realizar estos cálculos, podrá encontrar muchas probabilidades de Poisson directamente en la tabla E.7, que se reproduce de manera parcial en la tabla 5.5. La

tabla E.7 indica las probabilidades correspondientes a la variable aleatoria de Poisson para los valores de  $X = 0, 1, 2, \dots$ , de una selección de valores del parámetro  $\lambda$ . Para encontrar la probabilidad de que lleguen exactamente dos clientes durante un minuto dado, cuando la media de clientes que llegan por minuto es de 3.0, se lee la probabilidad correspondiente a la fila  $X = 2$  y la columna  $\lambda = 3.0$  de la tabla. El resultado es 0.2240, como muestra la tabla 5.5.

**TABLA 5.5**

Cálculo de la probabilidad de Poisson para  $\lambda = 3$ .

$X$	$\lambda$			
	2.1	2.2	....	3.0
0	.1225	.1108	....	.0498
1	.2572	.2438	....	.1494
2	.2700	.2681	....	.2240
3	.1890	.1966	....	.2240
4	.0992	.1082	....	.1680
5	.0417	.0476	....	.1008
6	.0146	.0174	....	.0504
7	.0044	.0055	....	.0216
8	.0011	.0015	....	.0081
9	.0003	.0004	....	.0027
10	.0001	.0001	....	.0008
11	.0000	.0000	....	.0002
12	.0000	.0000	....	.0001

Fuente: Tabla E.7.

También es factible calcular las probabilidades de Poisson que aparecen en la tabla E.7 utilizando Excel o Minitab. En la figura 5.5 se muestra una hoja de trabajo de Excel para la distribución de Poisson con  $\lambda = 3$ . En la figura 5.6 aparece el resultado en Minitab.

**FIGURA 5.5**

Hoja de trabajo de Excel para el cálculo de probabilidades de Poisson.

A	B	C	D	E
1	Customer Arrivals Analysis			
2				
3		Data		
4	Average/Expected number of successes:	3		
5				
6	Poisson Probabilities Table			
7	X	P(X)		
8	0	0.049787		=POISSON(A8,\$E\$4, FALSE)
9	1	0.149361		=POISSON(A9,\$E\$4, FALSE)
10	2	0.224042		=POISSON(A10,\$E\$4, FALSE)
11	3	0.224042		=POISSON(A11,\$E\$4, FALSE)
12	4	0.168031		=POISSON(A12,\$E\$4, FALSE)
13	5	0.100819		=POISSON(A13,\$E\$4, FALSE)
14	6	0.050409		=POISSON(A14,\$E\$4, FALSE)
15	7	0.021604		=POISSON(A15,\$E\$4, FALSE)
16	8	0.008102		=POISSON(A16,\$E\$4, FALSE)
17	9	0.002701		=POISSON(A17,\$E\$4, FALSE)
18	10	0.000810		=POISSON(A18,\$E\$4, FALSE)
19	11	0.000221		=POISSON(A19,\$E\$4, FALSE)
20	12	0.000055		=POISSON(A20,\$E\$4, FALSE)
21	13	0.000013		=POISSON(A21,\$E\$4, FALSE)
22	14	0.000003		=POISSON(A22,\$E\$4, FALSE)
23	15	0.000001		=POISSON(A23,\$E\$4, FALSE)
24	16	0.000000		=POISSON(A24,\$E\$4, FALSE)
25	17	0.000000		=POISSON(A25,\$E\$4, FALSE)
26	18	0.000000		=POISSON(A26,\$E\$4, FALSE)
27	19	0.000000		=POISSON(A27,\$E\$4, FALSE)
28	20	0.000000		=POISSON(A28,\$E\$4, FALSE)

**FIGURA 5.6**

Cálculo de la distribución de Poisson para  $\lambda = 3$  en Minitab.

**Poisson with mean = 3**

x	P( X = x )
0	0.049787
1	0.149361
2	0.224042
3	0.224042
4	0.168031
5	0.100819
6	0.050409
7	0.021604
8	0.008102
9	0.002701
10	0.000810
11	0.000221
12	0.000055
13	0.000013
14	0.000003
15	0.000001

**EJEMPLO 5.5****CÁLCULO DE PROBABILIDADES DE POISSON**

Se sabe que el número de fallas mensuales que tienen las cajas de velocidades de los autobuses obedece a la distribución de Poisson, con una media de 2.5 fallas al mes. ¿Cuál es la probabilidad de que no se presenten fallas durante un mes determinado? Y de que se presente al menos una?

**SOLUCIÓN** Con la ecuación (5.8) de la página 167 con  $\lambda = 2.5$  (o utilizando la tabla E.7, Excel o Minitab), se sabe que la probabilidad de que no se presenten fallas durante un mes dado es:

$$P(X = 0) = \frac{e^{-2.5}(2.5)^0}{0!} = \frac{1}{(2.71828)^{2.5}(1)} = 0.0821$$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - 0.0821 \\ &= 0.9179 \end{aligned}$$

La probabilidad de que no se presentarán fallas durante un mes dado es de 0.0821. La probabilidad de que se presente al menos una falla es de 0.9179.

**PROBLEMAS PARA LA SECCIÓN 5.3****Aprendizaje básico**

- ASISTENCIA de PH Grade**
- 5.18** Suponga una distribución de Poisson.
- Si  $\lambda = 2.5$ , encuentre  $P(X = 2)$ .
  - Si  $\lambda = 8.0$ , encuentre  $P(X = 8)$ .
  - Si  $\lambda = 0.5$ , encuentre  $P(X = 1)$ .
  - Si  $\lambda = 3.7$ , encuentre  $P(X = 0)$ .

ASISTENCIA  
de PH Grade

**5.19** Suponga una distribución de Poisson.

- Si  $\lambda = 2.0$ , encuentre  $P(X \geq 2)$ .
- Si  $\lambda = 8.0$ , encuentre  $P(X \geq 3)$ .
- Si  $\lambda = 0.5$ , encuentre  $P(X \leq 1)$ .
- Si  $\lambda = 4.0$ , encuentre  $P(X \geq 1)$ .
- Si  $\lambda = 5.0$ , encuentre  $P(X \leq 3)$ .

**5.20** Suponga una distribución de Poisson con  $\lambda = 5.0$ . Determine la probabilidad de que

- $X = 1$ .
- $X < 1$ .
- $X > 1$ .
- $X \leq 1$ .

### Aplicación de conceptos

Puede resolver los problemas 5.21 a 5.31 manualmente, en Excel o en Minitab.

**5.21** Suponga que el número de errores que se presentan durante un día en una red de área local (LAN) se distribuye como una variable aleatoria de Poisson. La media de errores de red ocurridos durante un día es 2.4. ¿Cuál es la probabilidad de que en un día determinado:

- se presenten cero errores?
- se presente sólo un error?
- se presenten dos o más errores?
- se presenten menos de tres errores?

 **AUTO Examen** **5.22** El gerente de control de calidad de Marilyn's Cookies inspecciona un lote de galletas con chispas de chocolate que se acaban de preparar. Si el proceso de producción está bajo control, la media de chispas de chocolate por galleta es de 6.0. ¿Cuál es la probabilidad de que en cualquier galleta inspeccionada

- se encuentren menos de cinco chispas?
- se encuentren exactamente cinco chispas?
- se encuentren cinco o más chispas?
- se encuentren cuatro o cinco chispas?

**5.23** De acuerdo con los datos del problema 5.22, ¿cuántas galletas de un lote de 100 esperaría desechar el gerente, si las políticas de la empresa exigen que todas las galletas con chispas de chocolate deben tener al menos cuatro chispas?

**5.24** El Departamento de Transporte de EUA registra las estadísticas de las maletas maltratadas por cada 1,000 pasajeros. En 2003, Jet Blue tuvo 3.21 maletas maltratadas por cada 1,000 pasajeros. ¿Cuál es la probabilidad de que, con los próximos 1,000 pasajeros, Jet Blue tenga

- ninguna maleta maltratada?
- al menos una maleta maltratada?
- al menos dos maletas maltratadas?
- Compare los resultados de los incisos a) a c) con los de Delta en el problema 5.25, incisos a) a c).

**5.25** El Departamento de Transporte de EUA registra las estadísticas de las maletas maltratadas por cada 1,000 pasajeros. En 2003, Delta tuvo 3.84 maletas maltratadas por cada 1,000 pasajeros. ¿Cuál es la probabilidad de que, con los próximos 1,000 pasajeros, Delta tenga

- ninguna maleta maltratada?
- al menos una maleta maltratada?
- al menos dos maletas maltratadas?
- Compare los resultados de los incisos a) a c) con los de Jet Blue en el problema 5.24, incisos a) a c).

ASISTENCIA  
de PH Grade

**5.26** Con base en experiencias anteriores, se supone que el número de imperfecciones por pie en los rollos de papel con graduación 2 obedece a una distribución de Poisson, con una media de 1 imperfección por cada cinco pies de papel (0.2 imperfecciones por pie). ¿Cuál es la probabilidad de que

- en un rollo de un pie existan al menos dos imperfecciones?
- en un rollo de 12 pies exista al menos 1 imperfección?
- en un rollo de 50 pies existan entre 5 y 15 (inclusive) imperfecciones?

**5.27** J.D. Power y Asociados calculan y publican varias estadísticas relacionadas con la calidad de los automóviles. La calificación Calidad Inicial mide el número de problemas por cada automóvil nuevo que se vende. Con los modelos 2003, el Lexus fue el mejor con 1.63 problemas por automóvil. El Kia coreano fue el peor, con 5.09 problemas por automóvil (L. Hawkins, "Finding a Car That's Built to Last?" *The Wall Street Journal*, 9 de julio, 2003, D1, D5). Sea la variable aleatoria  $X$  igual al número de problemas de un Lexus recién comprado.

- ¿Cuáles suposiciones se deben hacer para distribuir  $X$  como una variable aleatoria de Poisson? ¿Son razonables tales suposiciones?

Haciendo las suposiciones mencionadas en el inciso a), si usted compró un Lexus 2003, ¿cuál es la probabilidad de que este automóvil nuevo:

- no presente problemas?
- presente dos o menos problemas?
- Elabore una definición operacional de "problema". ¿Por qué es importante una definición operacional al interpretar la calificación Calidad Inicial?

**5.28** Consulte los datos del problema 5.27. Si usted compró un Kia 2003, ¿cuál es la probabilidad de que este automóvil nuevo:

- no presente problemas?
- presente dos o menos problemas?
- Compare sus respuestas de los incisos a) y b) con las correspondientes al Lexus del problema 5.27, incisos b) y c).

**5.29** Durante 2004, tanto el Lexus como el Kia mejoraron su desempeño (D. Hakim, "Hyundai Near Top of a Quality Ranking", *The New York Times*, 29 de abril, 2004, C.8). El Lexus presentó 0.87 problemas por automóvil y el Kia coreano tuvo 1.53 problemas por automóvil. Si usted compró un Lexus 2004, ¿cuál es la probabilidad de que este automóvil nuevo:

- no presente problemas?
- presente dos o menos problemas?
- Compare sus respuestas de los incisos a) y b) con las correspondientes al Lexus 2003 del problema 5.27, incisos b) y c).

**5.30** Consulte los datos del problema 5.29. Si usted compró un Kia 2004, ¿cuál es la probabilidad de que este automóvil nuevo:

- no presente problemas?
- presente dos o menos problemas?
- Compare sus respuestas de los incisos a) y b) con las correspondientes al Kia 2003 del problema 5.28, incisos a) y b).

**5.31** En su empresa existe un número telefónico gratuito, a disposición de los clientes desde las 9 AM hasta las 9 PM, para registrar las quejas relacionadas con algún producto comprado

a la empresa. Los datos recabados antes señalan que se reciben 0.4 llamadas por minuto.

a. ¿Cuáles propiedades deben ser verdaderas con respecto a la situación antes descrita para utilizar una distribución de Poisson con la finalidad de calcular las probabilidades relacionadas con el número de llamadas telefónicas recibidas durante un periodo de 1 minuto?

Suponiendo que esta situación concuerda con las propiedades que usted analizó en el inciso a), ¿cuál es la probabilidad de que durante un periodo de 1 minuto:

- b. no se reciban llamadas telefónicas?
- c. se reciban tres o más llamadas telefónicas?
- d. ¿Cuál es el número máximo de llamadas telefónicas que se recibirán durante un periodo de 1 minuto el 99.99% del tiempo?

## RESUMEN

En este capítulo estudió la esperanza matemática y el desarrollo y aplicación de las distribuciones binomial y de Poisson. En el escenario “Uso de la estadística” aprendió a calcular probabilidades a partir de la distribución binomial relacionada con la observación de las facturas marcadas por el sistema de información contable utilizado por la empresa de remodelaciones Saxon. En el siguiente capítulo, se desarrollará la distribución continua más importante: la distribución normal.

Como ayuda al decidir cuál de las dos distribuciones de probabilidad debe utilizar para una situación particular, debe plantearse la siguiente pregunta:

- ¿Existe un número fijo de observaciones  $n$ , cada una de las cuales se clasifica como éxito o fracaso; o existe un área de oportunidad? Si hay un número fijo de observaciones  $n$ , cada una de las cuales se clasifica como éxito o fracaso, utilice la distribución binomial. Si hay un área de oportunidad, utilice la distribución de Poisson.

## FÓRMULAS IMPORTANTES

### Valor esperado $\mu$ de una variable aleatoria discreta

$$\mu = E(X) = \sum_{i=1}^N X_i P(X_i) \quad (5.1)$$

### Varianza de una variable aleatoria discreta

$$\sigma^2 = \sum_{i=1}^N [X_i - E(X)]^2 P(X_i) \quad (5.2)$$

### Desviación estándar de una variable aleatoria discreta

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^N [X_i - E(X)]^2 P(X_i)} \quad (5.3)$$

### Combinaciones

$${}_n C_X = \frac{n!}{X!(n-X)!} \quad (5.4)$$

### Distribución de probabilidad binomial

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \quad (5.5)$$

### Media de la distribución binomial

$$\mu = E(X) = np \quad (5.6)$$

### Desviación estándar de la distribución binomial

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{np(1-p)} \quad (5.7)$$

### Distribución de la probabilidad de Poisson

$$P(X) = \frac{e^{-\lambda} \lambda^X}{X!} \quad (5.8)$$

## CONCEPTOS CLAVE

área de oportunidad 166  
desviación estándar de una variable aleatoria discreta 156  
distribución de Poisson 166  
distribución de probabilidad binomial 158

distribución de probabilidad de una variable aleatoria discreta 154  
modelo matemático 158  
regla de combinaciones 159

valor esperado de  $\mu$  de una variable aleatoria discreta 155  
varianza de una variable aleatoria discreta 156

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**5.32** ¿Cuál es el significado del valor esperado de una distribución de probabilidad?

**5.33** ¿Cuáles son las cuatro propiedades que se requieren de una situación para utilizar la distribución binomial?

**5.34** ¿Cuáles son las cuatro propiedades que se requieren de una situación para utilizar la distribución de Poisson?

### Aplicación de conceptos

Puede resolver los problemas 5.35 a 5.51 manualmente, en Excel o en Minitab.

**5.35** Los seguros de contingencias permiten a los promotores de espectáculos deportivos y de entretenimiento contar con protección de pérdidas financieras causadas por circunstancias fuera de su control, como los aguaceros. Por ejemplo, cada primavera el Ayuntamiento Central de la ciudad de Cincinnati pone en marcha el Festival de Cincinnati. Esa época del año es lluviosa en la zona, y la posibilidad de que llueva más de una pulgada durante un fin de semana es de alrededor de una de cuatro. En un artículo publicado por el *Cincinnati Enquirer*, escrito por Jim Knippenberg (“Chicken Pox Means 3 Dog Night Remedy”, *Cincinnati Enquirer*, 28 de mayo, 1997, E1), se muestran los detalles de la póliza de seguro contratada por el Ayuntamiento Central. La póliza abarca el pago de \$100,000 dólares si llueve más de una pulgada durante el fin de semana del festival. Se informó que el costo de la póliza fue de \$6,500 dólares.

- Determine si cree o no que el monto de esas cifras es correcto. (Sugerencia: Calcule el valor esperado de las utilidades que recibirá la empresa aseguradora.)
- Suponga que los montos son correctos. ¿Esta póliza es una buena transacción para el Ayuntamiento Central de Cincinnati?

**5.36** Desde 1872 hasta 2000, los precios de las acciones subieron el 74% de los años (Mark Hulbert, “The Stock Market Must Rise in 2002? Think Again”, *The New York Times*, 6 de diciembre, 2001, Business, 6). Con base en esta información, y suponiendo una distribución binomial, ¿cuál cree que es la probabilidad de que aumenten los precios en el mercado de valores

- el próximo año?
- el año siguiente al próximo?
- en cuatro de los próximos cinco años?
- en ninguno de los próximos cinco años?
- Para este caso, ¿qué suposición de la distribución binomial no sería válida?

**5.37** El costo medio de una llamada telefónica controlada por un sistema automático de servicio al cliente es de \$0.45. El costo medio de una llamada telefónica a través de una operadora es de \$5.50. Sin embargo, conforme más empresas instrumentan sistemas automáticos, aumenta más la molestia de los clientes con ese sistema. Muchos clientes se apresuran a salir del sistema automático cuando reciben una opción como “Para hablar con un representante de ventas, marque cero”. De acuerdo con el Centro para Conservación de Clientes, el 40% de todas las

personas que hablan a los servicios automáticos de servicio al cliente seleccionan comunicarse con una operadora cuando cuentan con esa opción (Jane Spencer, “In Search of the Operator”, *The Wall Street Journal*, 8 de mayo, 2002, D1).

Si 10 clientes independientes entre sí llaman a un sistema automático de servicio al cliente, ¿cuál es la probabilidad de que

- ninguno seleccione automáticamente la opción para hablar con una operadora?
- exactamente uno seleccione la opción para hablar con una operadora?
- dos o menos seleccionen automáticamente la opción para hablar con una operadora?
- los 10 seleccionen de manera automática la opción para hablar con una operadora?
- Si los 10 seleccionan automáticamente la opción para hablar con una operadora, ¿cree que la cifra del 40% mencionada en el artículo es aplicable a este sistema en particular? Explique por qué.

**5.38** Una teoría referente al índice industrial Dow Jones afirma que es probable que éste se incremente durante los años de elección presidencial en Estados Unidos. De 1964 a 2000, el índice industrial Dow Jones ha aumentado en ocho de los 10 años en los que ha habido elección presidencial en ese país. Suponiendo que este indicador es un evento aleatorio sin valor de pronóstico, usted esperaría que fuese correcto el 50% de las veces. ¿Cuál es la probabilidad de que el índice industrial Dow Jones aumente en ocho o más de los 10 años de elección presidencial estadounidense, si la verdadera probabilidad de incremento es

- 0.50?
- 0.70?
- 0.90?
- Con base en los resultados de los incisos a) a c), ¿cuál cree que es la probabilidad de que se incremente el índice industrial Dow Jones en un año en el que hay elecciones presidenciales en Estados Unidos?

**5.39** Priority Mail es la opción que ofrece el servicio postal estadounidense como competencia para las empresas de mensajería como Federal Express. Un artículo publicado en el *Wall Street Journal* presentó interesantes conclusiones en las que se comparan los envíos realizados por medio de Priority Mail con los más económicos embarques en primera clase (Rick Brooks, “New Data Reveal ‘Priority Mail’ Is Slower Than a Stamp”, *The Wall Street Journal*, 29 de mayo, 2002, D1). Al comparar los envíos para entrega en 3 días, se supo que las entregas en primera clase no llegaron a tiempo el 19% de las veces, mientras que Priority Mail falló el 33% de las veces. Hay que hacer notar que a la fecha de publicación del artículo, el costo mínimo de los envíos en primera clase era de \$0.34 y el de Priority Mail era de \$3.50.

Si se envían 10 artículos en primera clase, a 10 destinos distintos que le aseguran que están en una ubicación de entrega en 3 días, ¿cuál es la probabilidad de que:

- ningún artículo tarde más de 3 días?

- b. un artículo tarde exactamente más de 3 días?
- c. dos o más artículos tarden más de tres días?
- d. ¿Cuáles son la media y la desviación estándar de la distribución de probabilidad?

- 5.40** Consulte el problema 5.39. Si los envíos se realizan utilizando Priority Mail, ¿cuál es la probabilidad de que
- a. ningún artículo tarde más de 3 días?
  - b. un artículo tarde exactamente más de 3 días?
  - c. dos o más artículos tarden más de 3 días?
  - d. ¿Cuáles son la media y la desviación estándar de la distribución de probabilidad?
  - e. Compare los resultados de los incisos a) a c) con los del problema 5.39, incisos a) a c).

**5.41** La publicidad en las salas de cine está aumentando. Por lo general, de 60 a 90 segundos de duración, estos anuncios son más largos y extravagantes, y muestran la tendencia a sostener más audiencias cautivas que los anuncios de televisión. Así pues, no resulta sorprendente que las tasas de retención de los espectadores de los anuncios en cine sean superiores que las de los anuncios por televisión. De acuerdo con una encuesta realizada por la división ComQUEST del BBM Bureau of Measurement de Toronto, la probabilidad de que un espectador recuerde un anuncio que vio en el cine es de 0.74, mientras que la probabilidad de que un espectador recuerde un anuncio de 30 segundos que vio por televisión es de 0.37 (Nate Hendley, "Cinema Advertising Comes of Age", *Marketing Magazine*, 6 de mayo, 2002, 16).

- a. La probabilidad de 0.74 reportada por el BBM Bureau of Measurement, ¿quedaría mejor clasificada como probabilidad clásica *a priori*, probabilidad clásica empírica o probabilidad subjetiva?
- b. Suponga que se seleccionan como muestra, de manera aleatoria, 10 espectadores de un anuncio en cine. Considere la variable aleatoria definida por el número de espectadores que recuerdan el anuncio. ¿Cuáles suposiciones se deben hacer, para asumir que la distribución de esta variable es la de una variable aleatoria binomial?
- c. Suponiendo que el número de espectadores que recuerdan el anuncio en el cine es una variable aleatoria binomial, ¿cuáles son la media y la desviación estándar de esta distribución?
- d. Con base en su respuesta al inciso c), si ninguno de los espectadores puede recordar el anuncio, ¿qué se infiere con respecto a la probabilidad de 0.74 mencionada en el artículo?

**5.42** Consulte el problema 5.41. Calcule la probabilidad de que de 10 espectadores:

- a. ninguno recuerde el anuncio.
- b. los 10 recuerde el anuncio.
- c. más de la mitad recuerde el anuncio.
- d. ocho o más recuerden el anuncio.

**5.43** Consulte el problema 5.42. Calcule la probabilidad de que, para un anuncio en televisión con la probabilidad de retención dada de 0.37, para 10 espectadores:

- a. ninguno recuerde el anuncio.
- b. los 10 recuerden el anuncio.
- c. más de la mitad recuerde el anuncio.
- d. ocho o más recuerden el anuncio.

- e. Compare los resultados de los incisos a) a d) con los del problema 5.42, incisos a) a d).

**5.44** En una encuesta realizada por el Council for Marketing and Opinion Research (CMOR), grupo comercial no lucrativo del ramo de la investigación con sede en Cincinnati, 1,628 de los 3,700 adultos con los que se pusieron en contacto en Estados Unidos se rehusan a participar en encuestas por teléfono (Steve Jarvis, "CMOR Finds Survey Refusal Rate Still Rising", *Marketing News*, 4 de febrero, 2002, 4). Suponga que llama de forma aleatoria a 10 adultos en Estados Unidos y les pide que participen en una encuesta telefónica. Utilizando los resultados del estudio realizado por CMOR, ¿cuál es la probabilidad de que:

- a. los 10 se rehúsen?
- b. se rehúsen exactamente 5?
- c. se rehúsen por lo menos 5?
- d. se rehúsen menos de 5?
- e. menos de 5 acepten participar en la encuesta?
- f. ¿Cuál es el número esperado de personas que se rehusará a participar? Explique el significado práctico de este número.

**5.45** Las empresas emisoras de tarjetas de crédito están aumentando sus ingresos por medio de elevar los intereses moratorios que cobran a sus clientes. De acuerdo con un estudio realizado por [cardweb.com](http://cardweb.com), los intereses [mocardweb.com](http://mocardweb.com), los intereses moratorios son la tercera mayor fuente de ingresos para las emisoras de tarjetas de crédito, después de los intereses comunes y las comisiones que se cobran a quienes aceptan sus tarjetas. En uno de los últimos años, el 58% de todos los tarjetahabientes tuvieron que pagar intereses moratorios alguna vez (Ron Lieber, "Credit-Card Firms Collect Record Levels of Late Fees", *The Wall Street Journal*, 21 de mayo, 2002, D1).

Si se selecciona una muestra aleatoria de 20 tarjetahabientes, ¿cuál es la probabilidad de que:

- a. ninguno tenga que pagar intereses moratorios?
- b. no más de 5 tengan que pagar intereses moratorios?
- c. más de 10 tengan que pagar intereses moratorios?
- d. ¿Qué suposiciones tuvo que hacer para responder a los incisos a) al c)?

**5.46** Para quienes se dedican al comercio electrónico, no basta con hacer que un cliente visite su sitio Web. Estos comerciantes también deben persuadir a los compradores en línea de que realicen una compra y gasten dinero. Los expertos de Andersen Consulting estiman que el 88% de los compradores en red abandonan sus carritos de compras virtuales antes de finalizar la transacción (Rebecca Quick, "The Lessons Learned", *The Wall Street Journal*, 17 de abril, 2000, R6). Considere una muestra de 20 clientes que visitan un sitio Web de comercio electrónico, y suponga que la probabilidad de que un cliente abandone el sitio antes de finalizar la transacción es de 0.88. Utilice el modelo binomial para responder a las siguientes preguntas:

- a. ¿Cuál es el valor esperado, o media, de la distribución binomial?
- b. ¿Cuál es la desviación estándar de la distribución binomial?
- c. ¿Cuál es la probabilidad de que los 20 clientes abandonen el sitio sin completar una transacción?
- d. ¿Cuál es la probabilidad de que 18 clientes o más abandonen el sitio sin completar una transacción?
- e. ¿Cuál es la probabilidad de que 15 clientes o más abandonen el sitio sin completar una transacción?

**5.47** Consulte el problema 5.46. Si el sitio Web se optimiza de tal manera que sólo el 70% de los clientes abandonen el sitio sin completar su transacción,

- ¿cuál es el valor esperado, o media, de la distribución binomial?
- ¿cuál es la desviación estándar de la distribución binomial?
- ¿cuál es la probabilidad de que los 20 clientes abandonen el sitio sin completar una transacción?
- ¿cuál es la probabilidad de que 18 clientes o más abandonen el sitio sin completar una transacción?
- ¿cuál es la probabilidad de que 15 clientes o más abandonen el sitio sin completar una transacción?
- Compare los resultados de los incisos a) a e) con los del problema 5.46, incisos a) a e).

**5.48** Una teoría referente al índice Standard & Poor's 500 asegura que si éste aumenta durante los primeros cinco días hábiles del año, es probable que aumente durante todo el año. De 1950 a 2003, el índice Standard & Poor's 500 registró incremento en sus índices durante los primeros días de 34 de esos años. Suponiendo que este indicador es un evento aleatorio sin valor de pronóstico, usted esperaría que fuese correcto el 50% de las veces. ¿Cuál es la probabilidad de que el índice Standard & Poor's 500 aumente en 29 años, si su verdadera probabilidad de aumento es

- 0.50?
- 0.70?
- 0.90?
- Con base en los resultados de los incisos a) a c), ¿cuál cree que es la probabilidad de que aumente el índice Standard & Poor's 500, si registra utilidades durante los primeros cinco días hábiles del año? Explique por qué.

**5.49** El término falsa correlación se aplica a la aparente relación que existe entre variables que no tienen una relación real o que se relacionan con otras variables que no se midieron. Un indicador del mercado de valores muy difundido en Estados Unidos y que es un ejemplo de falsa correlación es la relación que existe entre el ganador del Súper Bowl y el desempeño del índice industrial Dow Jones en ese año. El indicador establece que cuando el ganador del Súper Bowl es de la Conferencia Nacional, el índice industrial Dow Jones aumenta durante ese año.

Cuando el ganador del Súper Bowl es de la Conferencia Americana, el índice industrial Dow Jones desciende durante ese año. En el lapso comprendido entre 1967 y 2003, un periodo de 37 años, el indicador ha resultado correcto en 31 de ellos. Suponiendo que este indicador es un evento aleatorio sin valor de pronóstico, usted esperaría que fuese correcto el 50% de las veces.

- ¿Cuál es la probabilidad de que el indicador resulte correcto 31 veces o más en 37 años?
- ¿Qué le dice lo anterior acerca de la utilidad de este indicador?

**5.50** Las ventas de pelotas de golf en todo el mundo suman más de mil millones de dólares al año. La venta de gran número de pelotas se debe en gran parte a que los golfistas las pierden con un promedio de 4.5 pelotas por ronda de 18 hoyos ("Snapshots", [usatoday.com](http://usatoday.com), 29 de enero, 2004). Suponga que el número de pelotas de golf extraviadas en una ronda de 18 hoyos tiene una distribución de variable aleatoria de Poisson.

- ¿Qué suposiciones es necesario hacer para que el número de pelotas de golf extraviadas en una ronda de 18 hoyos tenga una distribución de variable aleatoria de Poisson? Haciendo la suposiciones mencionadas en el inciso a), ¿cuál es la probabilidad de que:
- se extravíen 0 pelotas en una ronda de 18 hoyos?
- se extravíen 5 pelotas o menos en una ronda de 18 hoyos?
- se extravíen 6 pelotas o más en una ronda de 18 hoyos?

**5.51** Un estudio realizado en las páginas de inicio de los sitios Web de las empresas incluidas en la lista Fortune 500 encontró que el promedio de enlaces inservibles por página de inicio es de 0.4 y que el promedio de errores ortográficos por página de inicio es de 0.16 (Nabil Tamimi, Murii Rajan y Rose Sebastianella, "Benchmarking the Home Pages of 'Fortune' 500 Companies", *Quality Progress*, julio de 2000). Utilice la distribución de Poisson para buscar la probabilidad de que una página de inicio seleccionada al azar tenga:

- exactamente 0 enlaces inservibles.
- 5 o más enlaces inservibles.
- exactamente 0 errores ortográficos.
- 10 o más errores ortográficos.

## CASO ACTUAL

### ADMINISTRACIÓN DEL SPRINGVILLE HERALD

El departamento de mercadotecnia del *Herald* trata de incrementar sus ventas por suscripción, a través de una intensa campaña de marketing directo que incluye mensajes por correo, cupones de descuento y solicitudes por teléfono. La retroalimentación procedente de estas actividades señala que, tanto para suscriptores como para clientes potenciales, la entrega del periódico temprano por la mañana es un factor muy importante.

Luego de varias sesiones con lluvia de ideas, un equipo compuesto por miembros de los departamentos de marketing y

de circulación resolvió que la garantía de entrega del periódico a una hora específica sería un buen argumento de venta para conservar suscriptores y conseguir nuevos. Este equipo concluyó que el *Herald* ofrecería a sus clientes la garantía de que recibirían su periódico a cierta hora o, de lo contrario, ese ejemplar sería gratuito.

Para ayudar al equipo a establecer una hora de entrega garantizada, Al Leslie, director de investigación, observó que el departamento de circulación tenía los datos que mostrarían ca-

da cuarto de hora el porcentaje de periódicos que faltaban por entregar, de 6:00 AM a 8:00 AM. Jan Shapiro recordó que en los formatos de suscripción se preguntaba a los clientes a qué hora esperaban recibir su ejemplar del *Herald*. Esos datos se combinaron después y se colocaron en una página Web interna del *Herald* (vea el archivo **Circulation\_Data.htm** que se encuentra en la carpeta HeraldCase del CD-ROM que acompaña este volumen o entre a [www.prenhall.com/HeraldCase/Circulation\\_Data.htm](http://www.prenhall.com/HeraldCase/Circulation_Data.htm)).

## EJERCICIOS

Revise los datos internos y proponga una hora razonable (al cuarto de hora más cercano) para garantizar la entrega. Como ayuda para explorar los efectos de su elección, calcule las siguientes probabilidades:

- SH5.1 Si en un día determinado selecciona una muestra de 50 clientes, ¿cuál es la probabilidad de que, dada su hora de entrega seleccionada:
- menos de tres clientes reciban un ejemplar gratuito?
  - 2, 3 o 4 clientes reciban un ejemplar gratuito?
  - más de 5 clientes reciban un ejemplar gratuito?
- SH5.2 Considere que los efectos de corrección del proceso de entrega del periódico provocan que el porcentaje de periódicos que no se entregan a la hora garantizada se reduzca en un 2%. Si en un día determinado se selecciona una muestra de 50 clientes, ¿cuál es la probabilidad de que, dada su hora de entrega seleccionada (y la mejoría en la entrega):
- menos de 3 clientes reciban un ejemplar gratuito?
  - 2, 3 o 4 clientes reciban un ejemplar gratuito?
  - más de 5 clientes reciban un ejemplar gratuito?

## REFERENCIAS

- Bernstein, P. L., *Against the Gods: The Remarkable Story of Risk* (Nueva York: Wiley, 1996).
- Emery, D. R. y J. D. Finnerty, *Corporate Financial Management*, 2a. ed. (Upper Saddle River, NJ: Prentice Hall, 2000).
- Kirk, R. L., ed., *Statistical Issues: A Reader for the Behavioral Sciences* (Belmont, CA: Wadsworth, 1972).
- Levine, D. M., P. Ramsey y R. Smidt, *Applied Statistics for Engineers and Scientists Using Microsoft Excel and Minitab* (Upper Saddle River, NJ: Prentice Hall, 2001).
- Mescove, S.A., M. G. Simkin y A. Barganoff, *Core Concepts of Accounting Information Systems*, 7a. ed. (Nueva York: John Wiley, 2001).
- Microsoft Excel 2003* (Redmond, WA: Microsoft Corp., 2002).
- Minitab for Windows Version 14* (State College, PA: Minitab Inc., 2004).

# Apéndice 5 Uso de software para distribuciones de probabilidad discreta

## A5.1 EXCEL

### Para el valor esperado de una variable aleatoria discreta

Abra el archivo **Expected Value.xls**. Esta hoja de trabajo contiene las entradas de la tabla 5.1 del ejemplo sobre aprobación de hipotecas de la página 155, y utiliza las funciones SUM y SQRT (raíz cuadrada) para calcular estadísticos. Para adaptar esta hoja de trabajo a otros problemas:

- Si tiene más o menos de siete resultados, primero añada o elimine filas a la tabla, seleccionando el rango de celdas A5:E5 y luego **Insertar → Celdas o Editar → Eliminar** (si aparece un cuadro de opciones, seleccione **Mover cel-**

**das hacia abajo** si está añadiendo celdas, o **Mover celdas hacia arriba** si las está eliminando).

- Si añade filas, copie las fórmulas del rango de celdas C4: E4 en las nuevas filas de la tabla.
- Introduzca una lista corregida de valores de  $X$  en la columna A, comenzando con 1 en la celda A5.
- Introduzca los nuevos valores de  $P(X)$  en la columna B.

### Para probabilidades binomiales

Abra el archivo de Excel **Binomial.xls**, que se muestra en la figura 5.2 de la página 162. Esta hoja de trabajo ya contiene las entradas del ejemplo de pedidos marcados utilizado en la sección 5.2. Esta hoja utiliza la función BINOMDIST para calcular

probabilidades binomiales (para mayor información, consulte la sección G.6).

Para adaptar esta hoja de trabajo a otros problemas:

- Si tiene más o menos de tres resultados, primero añada o elimine filas a la tabla, seleccionando **fila 15** y luego **Insertar → Celdas o Editar → Eliminar**.
- Si añade filas, copie las entradas del rango de celdas A14: B14 en toda la tabla, para actualizarla.
- Introduzca los valores de nuevo tamaño de la muestra y la probabilidad de éxito en las celdas B4 y B5.

O Vea la sección G.6 (**Binomial**) si desea que PHStat2 elabore esta hoja de trabajo para usted.

### Para probabilidades de Poisson

Abra el archivo de Excel **Poisson.xls**, que muestra la figura 5.5 de la página 168. Esta hoja de trabajo ya contiene las entradas del ejemplo sobre llegada de clientes utilizado en la sección 5.3. Esta hoja utiliza la función POISSON para calcular probabilidades de Poisson (para mayor información, consulte la sección G.7). Para adaptar esta hoja de trabajo a otros problemas:

- Si necesita más de 20 resultados, primero añada filas a la tabla, seleccionando **fila 9** y luego **Insertar → Filas**. Despues, copie las entradas del rango de celdas A8:B8 en toda la tabla, para actualizarla.
- Introduzca el **valor del número de éxitos promedio/esperado** en la celda E4.

O Vea la sección G.7 (**Poisson**) si desea que PHStat2 elabore esta hoja de trabajo para usted.

## A5.2 MINITAB

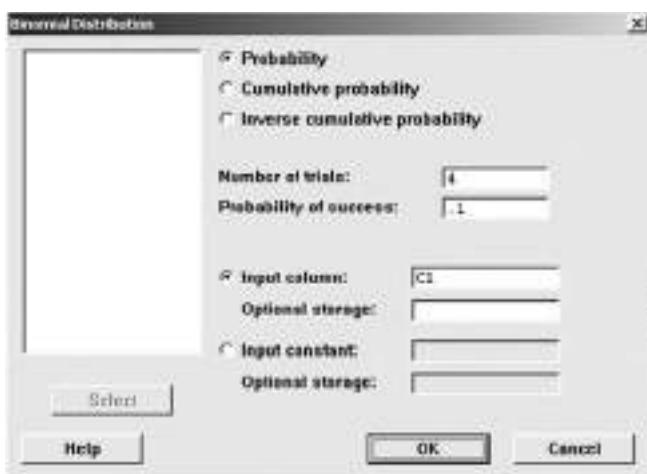
### Uso de Minitab para calcular probabilidades binomiales

Para ilustrar el uso de Minitab, tome en cuenta el sistema de información contable analizado en la sección 5.2. Para calcular los resultados de la figura 5.3 de la página 162,

1. Introduzca los valores **0, 1, 2, 3 y 4** en las filas 1 a 5 de la columna C1.
2. Seleccione **Calc → Probability Distributions → Binomial** para calcular probabilidades binomiales. En el cuadro de diálogo Binomial Distribution (distribución binomial, vea la figura A5.1), seleccione la opción **Probability** para calcular las probabilidades de  $X$  éxitos para todos los valores de  $X$ . En el cuadro de edición Number of trials: introduzca un tamaño de la muestra de 4. En el cuadro de editar Probability of success: introduzca **.10**. Seleccione el botón con la opción **Input column:** e introduzca **C1** en el cuadro editable. Dé clic en el botón **OK**.

### Uso de Minitab para calcular probabilidades de Poisson

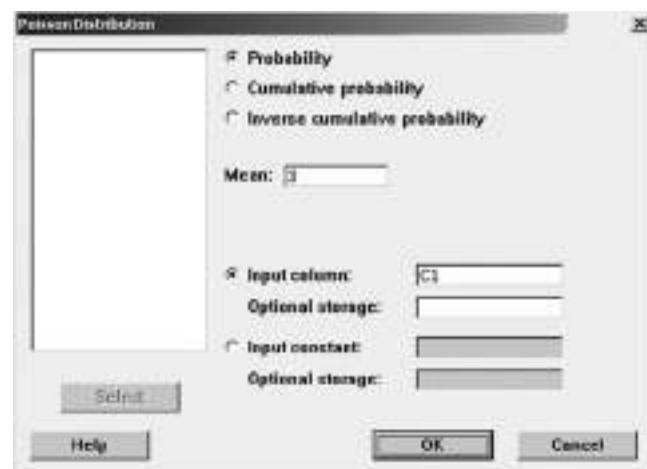
Para ilustrar la manera de calcular probabilidades de Poisson utilizando Minitab, volvamos al ejemplo del número de clientes



**FIGURA A5.1** Cuadro de diálogo Binomial Distribution en Minitab.

que llegan al banco, utilizado en la sección 5.3. Para calcular los resultados que se muestran en la figura 5.6 de la página 169,

1. Introduzca los valores **0 al 15** en las filas 1 a 16 de la columna C1.
2. Para calcular probabilidades de Poisson, seleccione **Calc → Probability Distributions → Poisson**. En el cuadro de diálogo Poisson Distribution (distribución de Poisson, vea la figura A5.2), seleccione la opción **Probability** para calcular las probabilidades de  $X$  éxitos para todos los valores de  $X$ . En el cuadro de edición Mean: introduzca el valor **λ de 3**. Seleccione el botón con la opción **Input column:** e introduzca **C1** en el cuadro de edición. Dé clic en **OK**.



**FIGURA A5.2** Ventana de diálogo Poisson Distribution de Minitab.

# CAPÍTULO 6

## La distribución normal

---

**USO DE LA ESTADÍSTICA:** Tiempo de descarga para la página principal de un sitio Web

**6.1 DISTRIBUCIONES DE PROBABILIDAD CONTINUA**

**6.2 LA DISTRIBUCIÓN NORMAL**

**6.3 EVALUACIÓN DE LA NORMALIDAD**

Evaluación de las propiedades

Construcción de un plano de probabilidad normal

**A.6 USO DEL SOFTWARE CON LA DISTRIBUCIÓN NORMAL**

*A6.1 Excel*

*A6.2 Minitab*

*A6.3 (Tema del CD-ROM) SPSS*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- A calcular probabilidades de una distribución normal
- A usar el plano de distribución normal para determinar si un conjunto de datos está distribuido de forma aproximadamente normal

# USO DE LA ESTADÍSTICA



## Tiempo de descarga para la página principal de un sitio Web

Suponga que usted es el diseñador del sitio Web OnCampus! que se dirige a los estudiantes universitarios. Para atraer y retener a los usuarios, necesita asegurarse de que la página principal se descargue rápidamente. Tanto el diseño de la página principal como la carga en el servidor Web de la empresa afectan el tiempo de descarga. Para revisar qué tan rápido se carga la página principal, se abre un navegador en una PC de las oficinas corporativas de OnCampus! y se mide el tiempo de descarga, cuántos segundos pasan desde que se enlaza con el sitio Web hasta que la página principal se abre por completo. Datos anteriores indican que la media del tiempo de descarga es de 7 segundos y la desviación estándar es de 2 segundos. Aproximadamente dos terceras partes del tiempo de descarga están entre 5 y 9 segundos, y cerca del 95% de los tiempos de descarga están entre los 3 y los 11 segundos. En otras palabras, los tiempos de descarga se distribuyen como una curva en forma de campana con un agrupamiento alrededor de la media de 7 segundos. ¿Cómo usaría esta información para responder las preguntas acerca de los tiempos de descarga de la página principal actual?

**E**n el capítulo anterior, los gerentes de Saxon Home Improvement Company, querían ser capaces de resolver problemas referentes al número de ocurrencias de cierto tipo de resultados en un tamaño de muestra dado. Como diseñador Web de OnCampus!, usted se enfrenta a una tarea diferente, una que implica una medición continua porque el tiempo de descarga podría tener cualquier valor y no sólo un número entero. ¿Cómo podría responder a preguntas como las siguientes acerca de esta *variable numérica continua*?

- ¿Qué proporción de las descargas de la página principal toman más de 10 segundos?
- ¿Cuántos segundos transcurren antes de que el 10% de las descargas se hayan realizado de forma completa?
- ¿Cuántos segundos transcurren antes de que el 99% de las descargas se hayan realizado por completo?
- ¿Cómo afectaría a las respuestas anteriores el rediseño de la página para que se descargue más rápido?

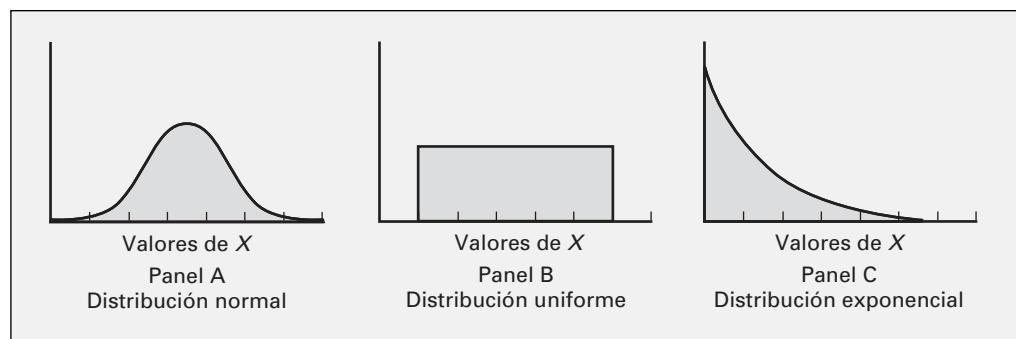
Como en el capítulo anterior, se utilizará la distribución de probabilidad como modelo. Leer este capítulo le ayudará a aprender acerca de las características de la distribución de la probabilidad continua y cómo usar la distribución normal para resolver problemas de negocios.

## 6.1 DISTRIBUCIONES DE PROBABILIDAD CONTINUA

Una **función de densidad de probabilidad continua** es una expresión matemática que define la distribución de los valores para una variable aleatoria continua. La figura 6.1 muestra gráficamente tres funciones de densidad de probabilidad continua. El panel A representa una distribución normal. La distribución normal es simétrica y con forma de campana, lo que implica que la mayoría de los valores tienden a agruparse alrededor de la media, la cual, por su forma simétrica, es igual a la mediana. Aunque desde el punto de vista teórico los valores en una distribución normal pueden clasificarse del infinito negativo al infinito positivo, la forma de la distribución hace poco posible que ocurran valores extremadamente grandes o muy pequeños. El panel B representa una distribución uniforme donde la probabilidad de ocurrencia de un valor es igualmente posible de ocurrir en cualquier lugar entre el menor valor  $a$  y el mayor valor  $b$ . La distribución uniforme, en ocasiones llamada distribución rectangular, es simétrica, y por lo tanto, la media es igual a la mediana. El panel C representa una distribución exponencial. Esta distribución está sesgada hacia la derecha, haciendo que la media sea más grande que la mediana. El rango de una distribución exponencial es de cero a infinito positivo, pero su forma hace que la ocurrencia de valores extremadamente grandes sea muy poco posible.

**FIGURA 6.1**

Tres distribuciones continuas.



## 6.2 LA DISTRIBUCIÓN NORMAL

La **distribución normal** (en ocasiones llamada distribución *gaussiana*) es la distribución continua que se utiliza más comúnmente en estadística. La distribución normal es de vital importancia en estadística por tres razones principales:

- Muchas variables continuas comunes en el mundo de los negocios tienen distribuciones que se asemejan estrechamente a la distribución normal.
- La distribución normal sirve para acercarse a diversas distribuciones de probabilidad discreta, como la distribución binomial y la distribución de Poisson.
- La distribución normal proporciona la base para la *estadística inferencial clásica* por su relación con el *teorema de límite central* (que se estudiará en el apartado 7.2).

La distribución normal se representa por la clásica forma de campana, ilustrada en el panel A de la figura 6.1. En la distribución normal, uno puede calcular la probabilidad de que varios valores ocurran dentro de ciertos rangos o intervalos. Sin embargo, la probabilidad *exacta* de un *valor particular* dentro de una distribución continua, como la distribución normal, es cero. Esta propiedad distingue a las variables continuas, que son medidas, de las variables discretas, las cuales son contadas. Como ejemplo, el tiempo (en segundos) se mide y no se cuenta. Por lo tanto, es factible determinar la probabilidad de que el tiempo de descarga para una página principal en un navegador de la Web esté entre 7 y 10 segundos o que la probabilidad de que el tiempo de descarga esté entre 8 y 9 segundos, o la probabilidad de que el tiempo de descarga esté entre 7.99 y 8.01 segundos. Sin embargo, la probabilidad de que el tiempo de descarga sea *exactamente* de 8 segundos es cero.

La distribución normal tiene importantes propiedades teóricas:

- Tiene una apariencia de forma de campana (y, por ende, es simétrica).
- Sus medidas de tendencia central (media, mediana y moda) son todas idénticas.
- Su “50% central” es igual a 1.33 desviaciones estándar. Esto significa que el rango intercuartil está contenido dentro de un intervalo de dos tercios de una desviación estándar por debajo de la media y de dos tercios de una desviación estándar por encima de la media.
- Su variable aleatoria asociada tiene un rango infinito ( $-\infty < X < \infty$ ).

En la práctica, muchas variables tienen distribuciones que se asemejan a las propiedades teóricas de la distribución normal. Los datos de la tabla 6.1 representan el espesor (en pulgadas) de 10,000 pulidores de cobre producidos por una gran empresa. La variable continua de interés, el espesor, puede aproximarse a la distribución normal. Las medidas del espesor de los 10,000 pulidores de cobre, se agrupan en el intervalo de 0.0190 a 0.0192 pulgadas y se distribuyen simétricamente alrededor de ese agrupamiento formando un patrón con forma de campana. Como se demuestra en la tabla 6.1, si este listado que no se traslapa (es *mutuamente excluyente*) contiene todos los intervalos de clase posibles (es *colectivamente exhaustivo*), las probabilidades sumarán 1. Tal distribución de probabilidad es una distribución de frecuencia relativa, como se describe en el apartado 2.3 donde, a excepción de dos clases abiertas al final, el punto medio de cada dos intervalos representa los datos en ese intervalo.

**TABLA 6.1**

Grosor de 10,000 pulidores de cobre.

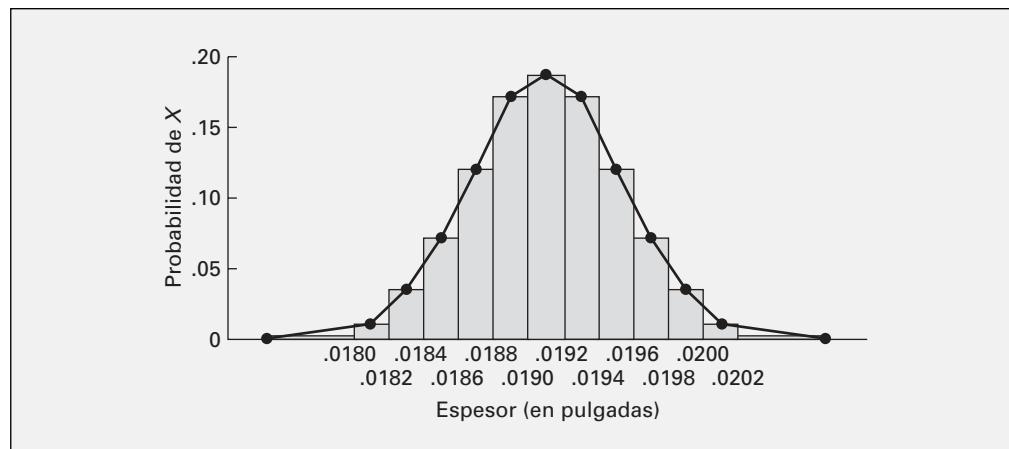
Grosor (en pulgadas)	Frecuencia relativa
Debajo de 0.0180	48/10,000 = 0.0048
0.0180 < 0.0182	122/10,000 = 0.0122
0.0182 < 0.0184	325/10,000 = 0.0325
0.0184 < 0.0186	695/10,000 = 0.0695
0.0186 < 0.0188	1,198/10,000 = 0.1198
0.0188 < 0.0190	1,664/10,000 = 0.1664
0.0190 < 0.0192	1,896/10,000 = 0.1896
0.0192 < 0.0194	1,664/10,000 = 0.1664
0.0194 < 0.0196	1,198/10,000 = 0.1198
0.0196 < 0.0198	695/10,000 = 0.0695
0.0198 < 0.0200	325/10,000 = 0.0325
0.0200 < 0.0202	122/10,000 = 0.0122
0.0202 o más	48/10,000 = 0.0048
Total	1.0000

La figura 6.2 representa el histograma de frecuencia relativa y el polígono para la distribución del espesor de 10,000 pulidores de cobre. Las tres propiedades teóricas de la distribución normal para estos datos, aproximadamente, se satisfacen; sin embargo, la cuarta no se sostiene. La variable aleatoria de interés, el espesor, no puede tener un valor de cero o menor, y un pulidor no puede ser tan espeso que se vuelva inútil. En la tabla 6.1 se ve que sólo 48 de cada 10,000 pulidores de cobre producidos tienen un espesor de 0.0202 pulgadas o más, mientras que se espera que el mismo número tenga un espesor por debajo de 0.0180 pulgadas. Por lo tanto, la posibilidad de obtener aleatoriamente un pulidor tan delgado o tan grueso es  $0.0048 + 0.0048 = 0.0096$ , o menos de 1 en 100.

**FIGURA 6.2**

Histograma de frecuencia relativa y polígono del espesor de 10,000 pulidores de cobre.

Fuente: Los datos fueron tomados de la tabla 6.1.



La expresión matemática que representa la función de densidad de probabilidad continua está indicada con el símbolo  $f(X)$ . Para la distribución normal, la **función de densidad de la probabilidad normal** se da en la ecuación (6.1).

#### LA FUNCIÓN DE DENSIDAD DE LA PROBABILIDAD NORMAL

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X-\mu)/\sigma]^2} \quad (6.1)$$

donde  $e$  es la constante matemática aproximada por 2.71828

$\pi$  es la constante matemática aproximada por 3.14159

$\mu$  es la media

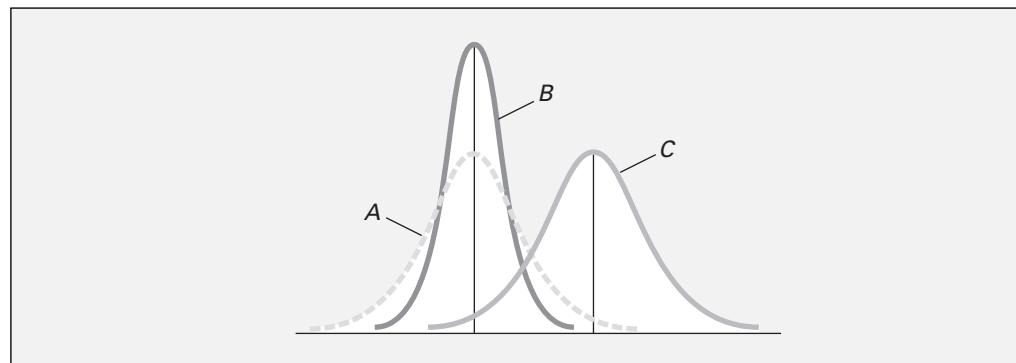
$\sigma$  es la desviación estándar

$X$  es cualquier valor de la variable continua, donde  $(-\infty < X < \infty)$

Puesto que  $e$  y  $\pi$  son constantes matemáticas, las probabilidades de una variable aleatoria  $X$  dependen sólo de dos parámetros de la distribución normal: la media  $\mu$  y la desviación estándar  $\sigma$ . Cada vez que se especifique una *combinación particular* de la  $\mu$  y la  $\sigma$ , se genera una distribución de probabilidad normal *diferente*. La figura 6.3 ilustra tres distribuciones normales diferentes. La distribución normal  $A$  y  $B$  tienen la misma media ( $\mu$ ), pero tienen desviaciones estándar diferentes. Las distribuciones  $A$  y  $C$  tienen la misma desviación estándar ( $\sigma$ ), pero tienen medias diferentes. Las distribuciones  $B$  y  $C$  representan dos funciones de densidad de probabilidad normal que difieren con respecto a la  $\mu$  y la  $\sigma$ .

**FIGURA 6.3**

Tres distribuciones normales.



La expresión matemática en la ecuación (6.1) es tediosa de calcular y requiere del cálculo integral. Por fortuna están disponibles tablas de probabilidad normal para evitar estos cálculos complicados. El primer paso para encontrar probabilidades normales es usar la **fórmula de transformación**, de la ecuación (6.2), para convertir cualquier variable aleatoria normal  $X$  en una **variable aleatoria normal estandarizada  $Z$** .

#### FÓRMULA DE TRANSFORMACIÓN

El valor  $Z$  es igual a la diferencia entre  $X$  y la media  $\mu$ , dividida por la desviación estándar  $\sigma$ .

$$Z = \frac{X - \mu}{\sigma} \quad (6.2)$$

Aunque los datos originales para la variable aleatoria  $X$  tenían una media  $\mu$  y una desviación estándar  $\sigma$ , la variable aleatoria estandarizada  $Z$  siempre tendrá una media  $\mu = 0$  y una desviación estándar  $\sigma = 1$ .

Al sustituir  $\mu = 0$  y  $\sigma = 1$  en la ecuación (6.1), la función de densidad de probabilidad de una variable normal estandarizada  $Z$  está dada en la ecuación (6.3).

#### LA FUNCIÓN DE DENSIDAD DE PROBABILIDAD NORMAL ESTANDARIZADA

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)Z^2} \quad (6.3)$$

Cualquier conjunto de valores distribuidos normalmente son susceptibles de convertirse a su forma estandarizada. Entonces se determinan las probabilidades deseadas usando la tabla E.2, la **distribución normal estandarizada acumulativa**.

Para ver cómo se aplica la fórmula de transformación y los resultados usados para encontrar las probabilidades de la tabla E.2,<sup>1</sup> recuerde que en el escenario “Uso de la estadística” de la página 178 los datos anteriores indicaban que el tiempo de descarga de la página Web se distribuye normalmente con una media  $\mu = 7$  segundos y una desviación estándar  $\sigma = 2$  segundos. En la figura 6.4 se observa que cada medición  $X$  tiene una medición estandarizada correspondiente  $Z$  calculada con la fórmula de transformación [ecuación (6.2)]. Por lo tanto, un tiempo de descarga de 9 segundos es equivalente a 1 unidad estandarizada (es decir, 1 desviación estándar por arriba de la media) porque

$$Z = \frac{9 - 7}{2} = +1$$

<sup>1</sup> Este texto utiliza la tabla E.2, la tabla normal estandarizada acumulativa. Al usar la tabla normal estandarizada, vea la tabla E.11 y la sección 6.1a del CD-ROM, “Using the Standardized Normal Distribution Table”.

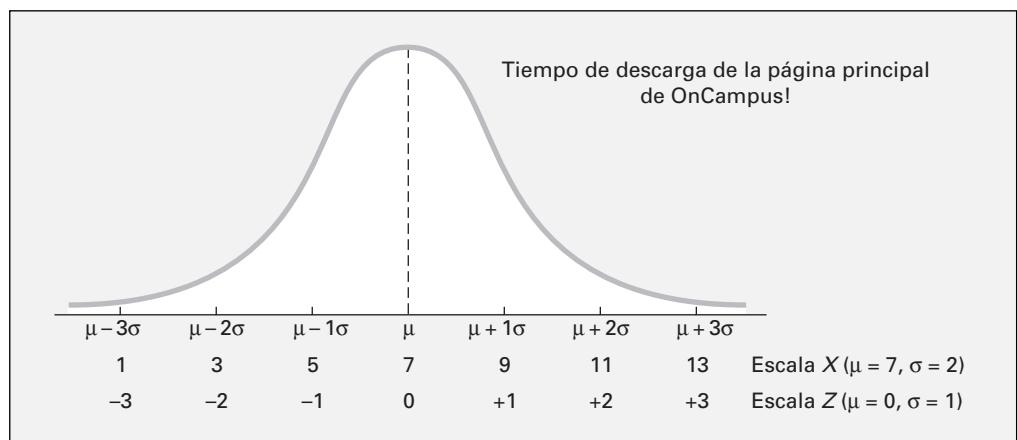
Un tiempo de descarga de 1 segundo es equivalente a 3 unidades estandarizadas (3 desviaciones estándar) por debajo de la media porque

$$Z = \frac{1 - 7}{2} = -3$$

Entonces, la desviación estándar es la unidad de medida. En otras palabras, un tiempo de 9 segundos es 2 segundos (es decir, 1 desviación estándar) más alto, o *más lento*, que la media de tiempo de 7 segundos. De forma similar, un tiempo de 1 segundo es 6 segundos (es decir, 3 desviaciones estándar) más bajo, o *más rápido*, que la media de tiempo.

**FIGURA 6.4**

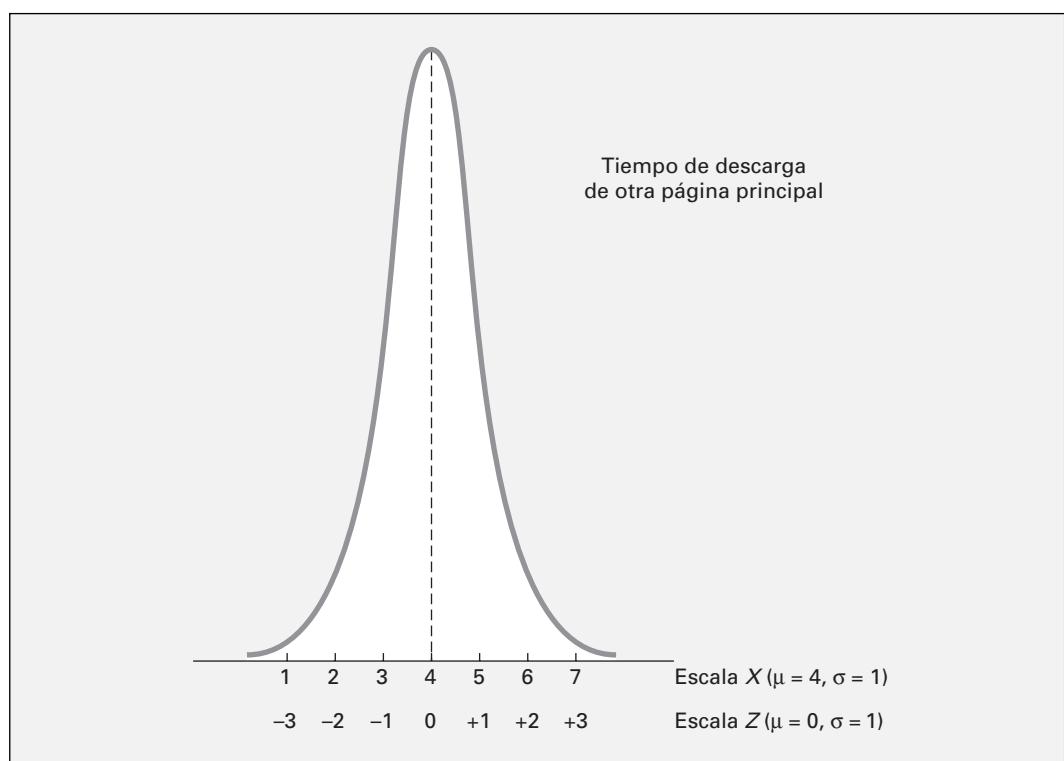
Escalas de transformación.



Para ilustrar mejor la fórmula de transformación, suponga que la página principal de otro sitio Web tiene un tiempo de descarga que se distribuye normalmente con una media  $\mu = 4$  segundos y una desviación estándar  $\sigma = 1$  segundo. Esta distribución se ilustra en la figura 6.5.

**FIGURA 6.5**

Una escala de transformación diferente.



Comparando estos resultados con los del sitio Web OnCampus!, se observa que el tiempo de descarga de 5 segundos es 1 desviación estándar por arriba de la media de tiempo de descarga porque

$$Z = \frac{5 - 4}{1} = +1$$

Un tiempo de 1 segundo es 3 desviaciones estándar por debajo de la media de tiempo de descarga porque

$$Z = \frac{1 - 4}{1} = -3$$

Las dos curvas con forma de campana en las figuras 6.4 y 6.5 muestran los polígonos de frecuencia relativa de distribuciones normales representando el tiempo de descarga (en segundos) para los dos sitios Web. Como los tiempos de descarga representan la población completa, las *probabilidades* o proporción de área bajo la curva entera deben sumar 1.

Suponga que desea encontrar la probabilidad de que el tiempo de descarga para el sitio Web OnCampus! sea menor a 9 segundos. Primero, se utiliza la ecuación (6.2) de la página 181 para transformar  $X = 9$  a unidades estandarizadas  $Z$ . Puesto que  $X = 9$  es una desviación estándar por arriba de la media,  $Z = +1.00$ . En seguida se utiliza la tabla E.2 para encontrar el área acumulativa bajo la curva normal calculada menor que (es decir, a la izquierda de)  $Z = +1.00$ . Para leer la probabilidad o área bajo la curva menor que  $Z = +1.00$  se revisa la columna  $Z$  de la tabla E.2 hasta localizar el valor  $Z$  de interés (en décimos) en la fila  $Z$  de 1.0. A continuación lea a través de esta fila hasta que intersecte la columna que contiene el lugar del valor  $Z$  en los centésimos. Por lo tanto, en el cuerpo de la tabla, la probabilidad tabulada para  $Z = 1.00$  corresponde con la intersección de la fila  $Z = 1.0$  con la columna  $Z = .00$ , como se muestra en la tabla 6.2, la cual se extrajo de la tabla E.2. Esta probabilidad es de 0.8413. Como se ilustra en la figura 6.6, hay un 84.13% de posibilidad de que el tiempo de descarga sea menor a 9 segundos.

**TABLA 6.2**

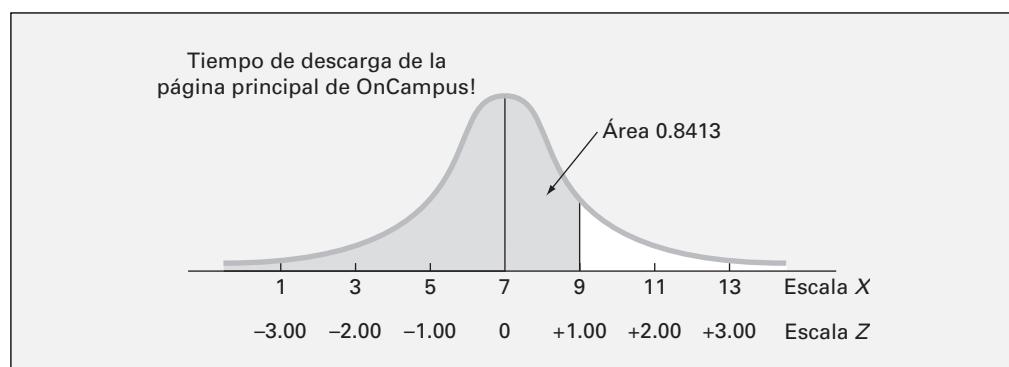
Encontrar el área acumulativa bajo la curva normal.

<b>Z</b>	<b>.00</b>	<b>.01</b>	<b>.02</b>	<b>.03</b>	<b>.04</b>	<b>.05</b>	<b>.06</b>	<b>.07</b>	<b>.08</b>	<b>.09</b>
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	<b>.8413</b>	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621

Fuente: Extraído de la tabla E.2.

**FIGURA 6.6**

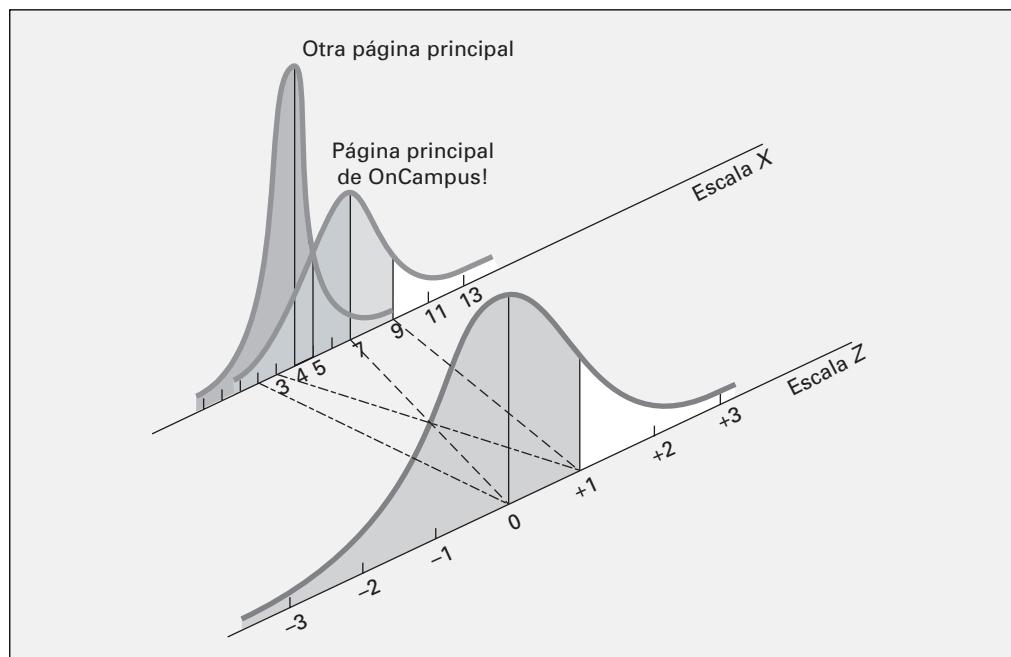
Determinación del área menor que  $Z$  de una distribución normal estandarizada acumulativa.



Sin embargo, para las otras páginas principales de la figura 6.5, en la página 182, se observa que un tiempo de 5 segundos es 1 unidad estandarizada por arriba de una media de tiempo de 4 segundos. Así, la probabilidad de que el tiempo de descarga sea menor a 5 segundos es también .08413. La figura 6.7 muestra que a pesar del valor de la media  $\mu$  y de la desviación estándar  $\sigma$  de una variable distribuida normalmente, la ecuación (6.2) permite transformar el problema a valores  $Z$ .

**FIGURA 6.7**

Demostración de la transformación de escalas para porciones correspondientes acumulativas bajo dos curvas normales.



Ahora que ha aprendido a usar la tabla E.2 con la ecuación (6.2), está en condiciones de responder muchas preguntas relacionadas con la página OnCampus! usando la distribución normal.

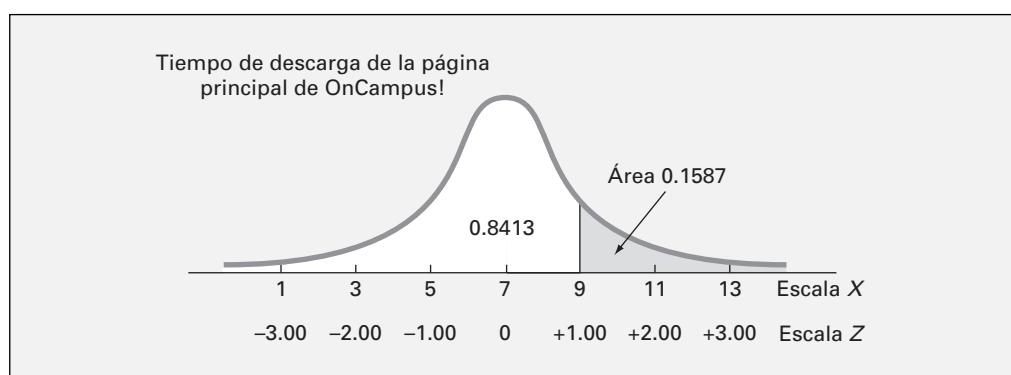
**EJEMPLO 6.1****ENCONTRAR  $P(X > 9)$** 

¿Cuál es la probabilidad de que el tiempo de descarga sea de más de 9 segundos?

**SOLUCIÓN** La probabilidad de que el tiempo de descarga sea menor a 9 segundos es de 0.8413 (vea la figura 6.6 en la página 183). Así, la probabilidad de que el tiempo de descarga sea mayor a 9 segundos es el *complemento* de que sea menor a 9 segundos,  $1 - 0.8413 = 0.1587$ . La figura 6.8 ilustra este resultado.

**FIGURA 6.8**

Encontrar  $P(X > 9)$ .



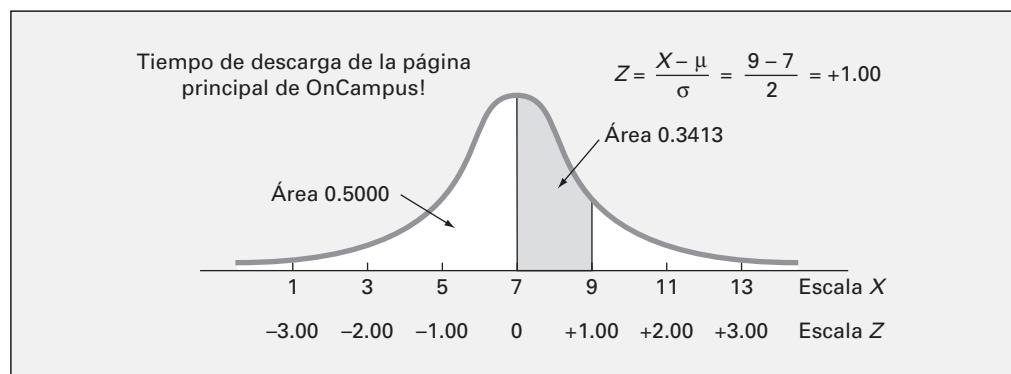
**EJEMPLO 6.2****ENCONTRAR  $P(7 < X < 9)$** 

¿Cuál es la probabilidad de que el tiempo de descarga esté entre 7 y 9 segundos?

**SOLUCIÓN** A partir de la figura 6.6 en la página 183, ya se ha determinado que la probabilidad de que el tiempo de descarga sea menor a 9 segundos es de 0.8413. Ahora deberá determinar la probabilidad de que el tiempo de descarga sea menor a 7 segundos y sustraer esto de la probabilidad de que el tiempo de descarga sea menor a 9 segundos, como se muestra en la figura 6.9.

**FIGURA 6.9**

Encontrar  $P(7 < X < 9)$ .



Usando la ecuación (6.2) vea la página 181

$$Z = \frac{7 - 7}{2} = 0.00$$

Al utilizar la tabla E.2, se sabe que el área bajo la curva normal menor que la media de  $Z = 0.00$  es 0.5000. Por lo tanto, el área bajo la curva entre  $Z = 0.00$  y  $Z = 1.00$  es  $0.8413 - 0.5000 = 0.3413$ .

**EJEMPLO 6.3****ENCONTRAR  $P(X < 7 \text{ o } X > 9)$** 

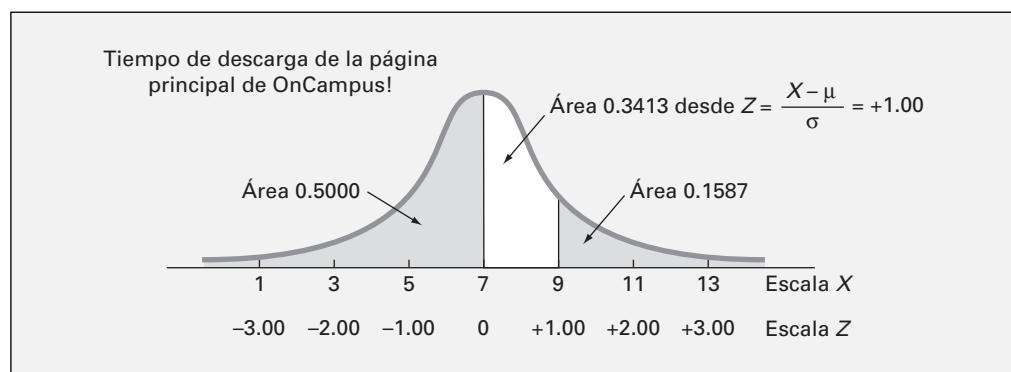
¿Cuál es la probabilidad de que el tiempo de descarga sea menor a 7 segundos o mayor a 9 segundos?

**SOLUCIÓN** De la figura 6.9, la probabilidad de que el tiempo de descarga sea entre 7 y 9 segundos es de 0.3413. La probabilidad de que el tiempo de descarga esté por debajo de los 7 segundos o por arriba de los nueve segundos es su complemento,  $1 - 0.3413 = 0.6587$ .

Otra manera de ver este problema, es calcular de forma separada la probabilidad de tiempo de descarga menor a 7 segundos y la probabilidad de tiempo de descarga mayor a 9 segundos, y después sumar estas dos probabilidades juntas para calcular el resultado deseado. Este resultado se representa en la figura 6.10. Puesto que la media y la mediana son la misma para datos distribuidos normalmente, el 50% de los tiempos de descarga están por debajo de 7 segundos. En el ejemplo 6.1, la probabilidad de tiempo de descarga de más de 9 segundos es de 0.1587. Por lo tanto, la probabilidad de que el tiempo de descarga esté por debajo de 7 o por arriba de 9 segundos,  $P(X < 7 \text{ o } X > 9)$ , es  $0.5000 + 0.1587 = 0.6587$ .

**FIGURA 6.10**

Encontrar  $P(X < 7 \text{ o } X > 9)$ .



**EJEMPLO 6.4****ENCONTRAR  $P(5 < X < 9)$** 

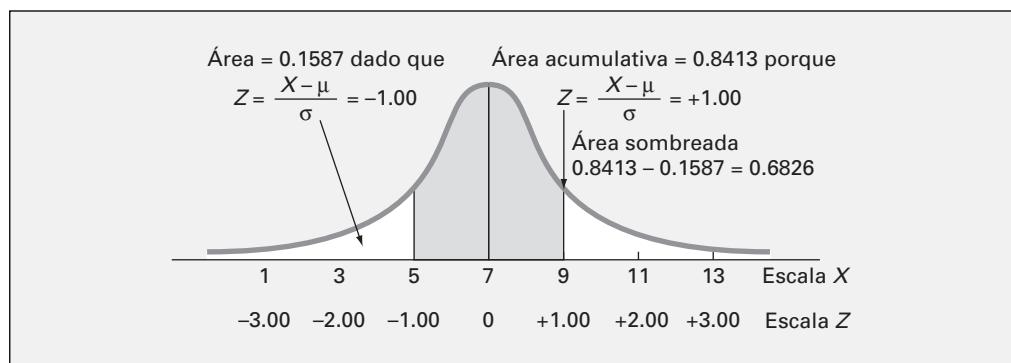
¿Cuál es la probabilidad de que el tiempo de descarga esté entre 5 y 9 segundos, es decir,  $P(5 < X < 9)$ ?

**SOLUCIÓN** En la figura 6.11 se observa que el área de interés se localiza entre dos valores, 5 y 9. Como la tabla E.2 le permite encontrar probabilidades menores que un valor en particular de interés, siga los siguientes tres pasos para encontrar la probabilidad deseada:

1. Determine la probabilidad de menos de 9 segundos.
2. Determine la probabilidad de menos de 5 segundos.
3. Sustraiga el resultado menor del mayor.

**FIGURA 6.11**

Encontrar  $P(5 < X < 9)$ .



Para este ejemplo, ya se completó el paso 1. El área bajo la curva normal menor que 9 segundos es 0.8413. Para encontrar el área bajo la curva normal menor que 5 segundos (paso 2),

$$Z = \frac{5 - 7}{2} = -1.00$$

En la tabla E.2, busque  $Z = -1.00$  y encuentre 0.1587. Para el paso 3, la probabilidad de que el tiempo de descarga sea entre 5 y 9 segundos es  $0.8413 - 0.1587 = 0.6826$ , como se muestra en la figura 6.11.

El resultado del ejemplo 6.4 es importante y le permite generalizar los resultados. Para cualquier distribución normal existe un 0.6826 de posibilidad de que un artículo seleccionado aleatoriamente caiga dentro de  $\pm 1$  desviación estándar de la media. A partir de la figura 6.12, se sabe que ligeramente más del 95% de los artículos caerán dentro de  $\pm 2$  desviaciones estándar. Por lo tanto, 95.44% de los tiempos de descarga están entre 3 y 11 segundos. A partir de la figura 6.13, se sabe que el 99.73% de los artículos caerán dentro de  $\pm 3$  desviaciones estándar por arriba o por debajo de la media. Así, el 99.73% de los tiempos de descarga están entre 1 y 13 segundos. Por lo tanto, es poco probable (0.0027, o sólo 27 en cada 10,000) que el tiempo de descarga sea tan rápido o tan lento que tome menos de 1 segundo o más de 13 segundos. Es por esto que  $6\sigma$  (es decir, de 3 desviaciones estándar por arriba de la media a 3 desviaciones estándar por debajo de la media) a menudo se usa como una *aproximación práctica del rango* para una distribución normal de datos.

Por lo tanto, para cualquier distribución normal:

- Aproximadamente el 68.26% de los artículos caerán dentro de  $\pm 1$  desviación estándar alrededor de la media.
- Aproximadamente el 95.44% de los artículos caerán dentro de  $\pm 2$  desviaciones estándar alrededor de la media.
- Aproximadamente el 99.73% de los artículos caerán dentro de  $\pm 3$  desviaciones estándar alrededor de la media.

El resultado anterior es la justificación para la regla empírica presentada en la página 96. Cuanto más cerca siga un conjunto de datos a la distribución normal, más precisa y exacta es la regla empírica.

**FIGURA 6.12**Encontrar  $P(3 < X < 11)$ .

El área por debajo es 0.0228 porque  
 $Z = \frac{X - \mu}{\sigma} = -2.00$

El área por debajo es 0.9772 porque  
 $Z = \frac{X - \mu}{\sigma} = +2.00$

**FIGURA 6.13**Encontrar  $P(1 < X < 13)$ .

Área por debajo es 0.00135 porque  
 $Z = \frac{X - \mu}{\sigma} = -3.00$

Área por debajo es 0.99865 porque  
 $Z = \frac{X - \mu}{\sigma} = +3.00$

**EJEMPLO 6.5****ENCONTRAR  $P(X < 3.5)$** 

¿Cuál es la probabilidad de que el tiempo de descarga esté por debajo de 3.5 segundos?

**SOLUCIÓN** Para calcular la probabilidad de que un tiempo de descarga esté por debajo de 3.5 segundos, es necesario examinar la región baja de la cola izquierda de la figura 6.14.

**FIGURA 6.14**Encontrar  $P(X < 3.5)$ .

El área es 0.0401 porque  $Z = \frac{X - \mu}{\sigma} = -1.75$

Para determinar el área bajo la curva por debajo de 3.5 segundos, primero calcule

$$Z = \frac{X - \mu}{\sigma} = \frac{3.5 - 7}{2} = -1.75$$

Busque el valor  $Z$  en la tabla E.2 apareando la fila  $Z$  apropiada (-1.7), con la columna  $Z$  apropiada (.05) como se muestra en la tabla 6.3 (que se extrajo de la tabla E.2). La probabilidad resultante o área bajo la curva menor que -1.75 desviaciones estándar por debajo de la media es 0.0401.

**TABLA 6.3**

Encontrar un área acumulativa bajo la curva normal.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
<b>-1.7</b>	<b>.0446</b>	<b>.0436</b>	<b>.0427</b>	<b>.0418</b>	<b>.0409</b>	<b>.0401</b>	<b>.0392</b>	<b>.0384</b>	<b>.0375</b>	<b>.0367</b>
<b>-1.6</b>	<b>.0548</b>	<b>.0537</b>	<b>.0526</b>	<b>.0516</b>	<b>.0505</b>	<b>.0495</b>	<b>.0485</b>	<b>.0475</b>	<b>.0465</b>	<b>.0455</b>

Fuente: Extraido de la tabla E.2.

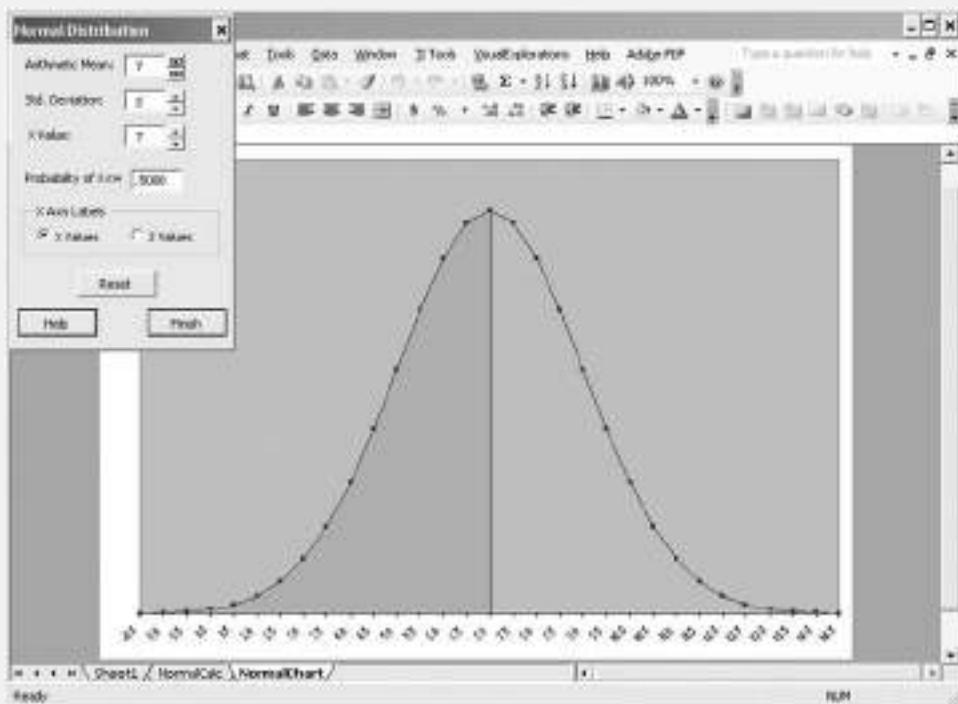
## EXPLORACIONES VISUALES Explorando la distribución normal

Utilice el comando de exploración visual de la distribución normal para ver los efectos de los cambios en la media y la desviación estándar en el área bajo una curva de distribución normal.

Abra el archivo **Visual Explorations.xla** y seleccione **VisualExplorations → Normal Distribution** de la barra del menú de Excel. Verá la curva normal para el ejemplo de "Uso de la estadística" de la descarga de la página principal, así como un panel de control flotante que le permite ajustar la forma de la curva y el área sombreada bajo la curva (vea ilustración abajo).

Use los botones del panel de control para cambiar los valores para la media, la desviación estándar y el valor  $X$ , mientras nota sus efectos en la probabilidad del valor  $X \leq$  y el área sombreada correspondiente bajo la curva (vea la ilustración de abajo). Si lo prefiere, seleccione la opción Valores Z para ver la curva normal rotulada con Valores Z.

Dé clic en el botón Reset para reajustar los valores del panel de control o dé clic en Ayuda para información adicional sobre el problema. Dé clic en Finalizar cuando haya terminado de explorar.



Los ejemplos 6.1 a 6.5 requieren que use las tablas normales para encontrar un área bajo la curva normal que corresponda a un valor  $X$  específico. En muchas circunstancias usted desea hacer lo contrario. Los ejemplos 6.6 y 6.7 ilustran cómo encontrar el valor  $X$  que corresponde a un área específica.

### EJEMPLO 6.6

#### ENCONTRAR EL VALOR $X$ PARA UNA PROBABILIDAD ACUMULATIVA DE 0.10

¿Cuánto tiempo transcurrirá (en segundos) antes de que el 10% de las descargas estén completas?

**SOLUCIÓN** Como se espera que el 10% de las páginas principales se descarguen por debajo de  $X$  segundos, el área bajo la curva normal menor a este valor  $Z$  es 0.1000. En el cuerpo de la tabla E.2, busque el área o probabilidad de 0.1000. El resultado más cercano es 0.1003, como se muestra en la tabla 6.4 (que se extrajo de la tabla E.2).

**TABLA 6.4**

Encontrar un Valor  $Z$  que corresponde a un área acumulativa particular (0.10) bajo la curva normal.

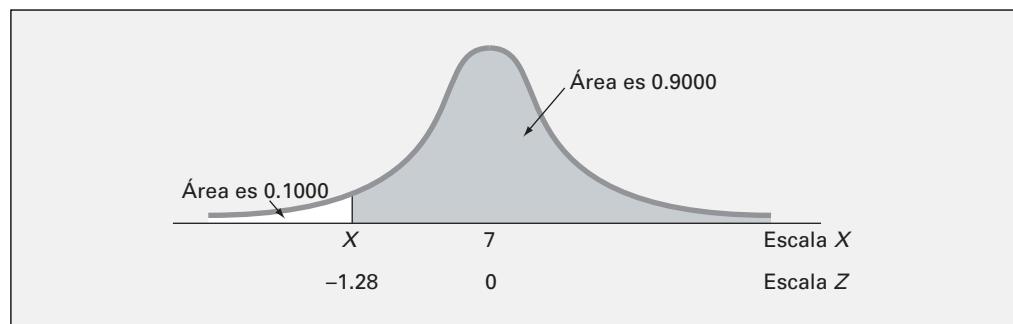
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	<b>.1151</b>	<b>.1131</b>	<b>.1112</b>	<b>.1093</b>	<b>.1075</b>	<b>.1056</b>	<b>.1038</b>	<b>.1020</b>	<b>.1003</b>	<b>.0985</b>

Fuente: Extraído de la tabla E.2.

Trabajando desde esta área a los márgenes de la tabla, el valor  $Z$  correspondiente a la fila  $Z$  (-1.2) y la columna  $Z$  (.08) es -1.28 (vea la figura 6.15).

**FIGURA 6.15**

Encontrar  $Z$  para determinar  $X$ .



Una vez que encuentre  $Z$ , use la fórmula de transformación de la ecuación (6.2) en la página 181 para determinar el valor  $X$  como sigue. Sea

$$Z = \frac{X - \mu}{\sigma}$$

entonces

$$X = \mu + Z\sigma$$

Al sustituir  $\mu = 7$ ,  $\sigma = 2$  y  $Z = -1.28$ ,

$$X = 7 + (-1.28)(2) = 4.44 \text{ segundos}$$

Por lo tanto, el 10% de los tiempos de descarga se realizan en 4.44 segundos o menos.

La ecuación (6.4) se utiliza para encontrar un valor  $X$ .

ENCONTRAR UN VALOR  $X$  ASOCIADO CON UNA PROBABILIDAD CONOCIDA  
El valor  $X$  es igual a la media  $\mu$  más el producto del valor  $Z$  y de la desviación estándar  $\sigma$ .

$$X = \mu + Z\sigma \quad (6.4)$$

Para encontrar un valor *particular* asociado con una probabilidad conocida, siga estos pasos.

1. Dibuje la curva normal, después coloque los valores para las medias en las escalas respectivas  $X$  y  $Z$ .
2. Encuentre el área acumulativa menor que  $X$ .
3. Sombree el área de interés.
4. Con la tabla E.2, determine el valor  $Z$  correspondiente al área bajo la curva normal por debajo de  $X$ .
5. Con la ecuación (6.4), resuelva para  $X$ :

$$X = \mu + Z\sigma$$

### EJEMPLO 6.7

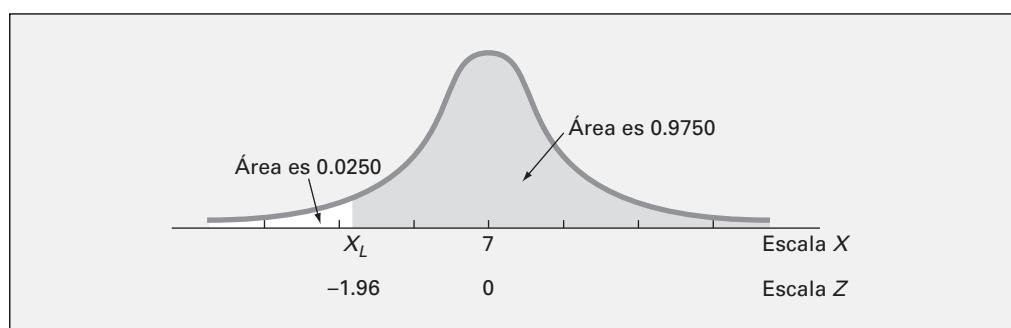
ENCONTRAR LOS VALORES  $X$  QUE INCLUYAN EL 95% DE LOS TIEMPOS DE DESCARGA

¿Cuáles son los valores inferiores y superiores de  $X$ , localizados simétricamente alrededor de la media, que incluyen el 95% de los tiempos de descarga?

**SOLUCIÓN** Primero, necesita encontrar el valor inferior de  $X$  (llamado  $X_L$ ). Despues debe encontrar el valor superior de  $X$  (llamado  $X_U$ ). Como el 95% de los valores están entre  $X_L$  y  $X_U$ , y  $X_L$  y  $X_U$  están a la misma distancia de la media, el 2.5% de los valores están por debajo de  $X_L$  (vea la figura 6.16).

**FIGURA 6.16**

Encontrar  $Z$  para determinar  $X_L$ .



Aunque  $X_L$  no es conocido, se puede encontrar el  $Z$  correspondiente porque el área bajo la curva normal menor a este  $Z$  es 0.0250. En el cuerpo de la tabla 6.5, busque la probabilidad 0.0250.

**TABLA 6.5**

Encuentre el valor  $Z$  correspondiente al área acumulativa de 0.0250 bajo la curva normal.

$Z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0232	.0314	.0307	.0301	.0294

Fuente: Extraido de la tabla E.2.

Trabajando desde el cuerpo de la tabla hacia los márgenes, se observa que el valor  $Z$  que corresponde a una fila  $Z$  en especial ( $-1.9$ ) y a una columna  $Z$  ( $.06$ ) es  $-1.96$ .

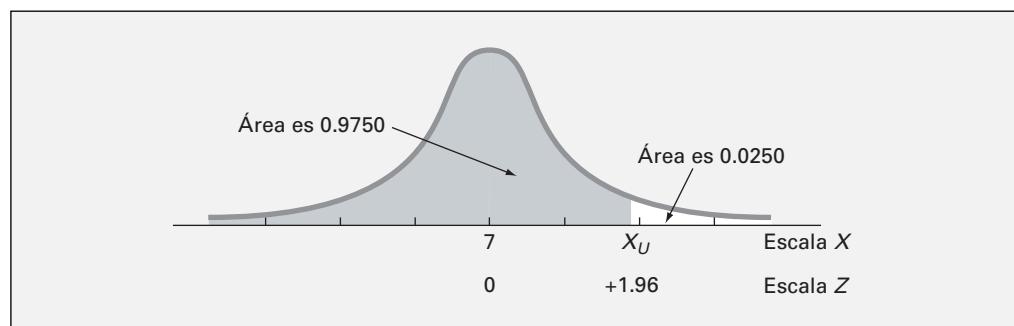
Una vez que encuentre  $Z$ , el último paso es usar la ecuación (6.4) de la página 190, de la siguiente manera,

$$\begin{aligned} X &= \mu + Z\sigma \\ &= 7 + (-1.96)(2) \\ &= 7 - 3.92 \\ &= 3.08 \text{ segundos} \end{aligned}$$

Se sigue un proceso parecido para encontrar  $X_U$ . Puesto que sólo el 2.5% de las descargas de las páginas principales toman más de  $X_U$  segundos, el 97.5% de las descargas de páginas principales toman menos de  $X_U$  segundos. Para la simetría de una distribución normal, el valor  $Z$  deseado, como se muestra en la figura 6.17, es  $+1.96$  (porque  $Z$  cae a la derecha de la media estandarizada de 0). También se puede extraer este valor  $Z$  de la tabla 6.6. Note que 0.9750 es el área bajo la curva normal menor que el valor  $Z$  de  $+1.96$ .

**FIGURA 6.17**

Encontrar  $Z$  para determinar  $X_U$ .



**TABLA 6.6**

Encontrar un valor  $Z$  que corresponda al área acumulativa de 0.975 bajo la curva normal.

$Z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
+1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
+1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
+2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

Fuente: Extraído de la tabla E.2.

Por lo tanto, al utilizar la ecuación (6.4) de la página 190,

$$\begin{aligned} X &= \mu + Z\sigma \\ &= 7 + (+1.96)(2) \\ &= 7 + 3.92 \\ &= 10.92 \text{ segundos} \end{aligned}$$

Por lo tanto, el 95% de los tiempos de descarga están entre 3.08 y 10.92 segundos.

También se podría utilizar Excel o Minitab para calcular probabilidades normales. La figura 6.18 ilustra una hoja de trabajo de Excel para los ejemplos 6.5 y 6.6, y la figura 6.19 ilustra un resultado de Minitab para los ejemplos 6.1 y 6.6.

**FIGURA 6.18**

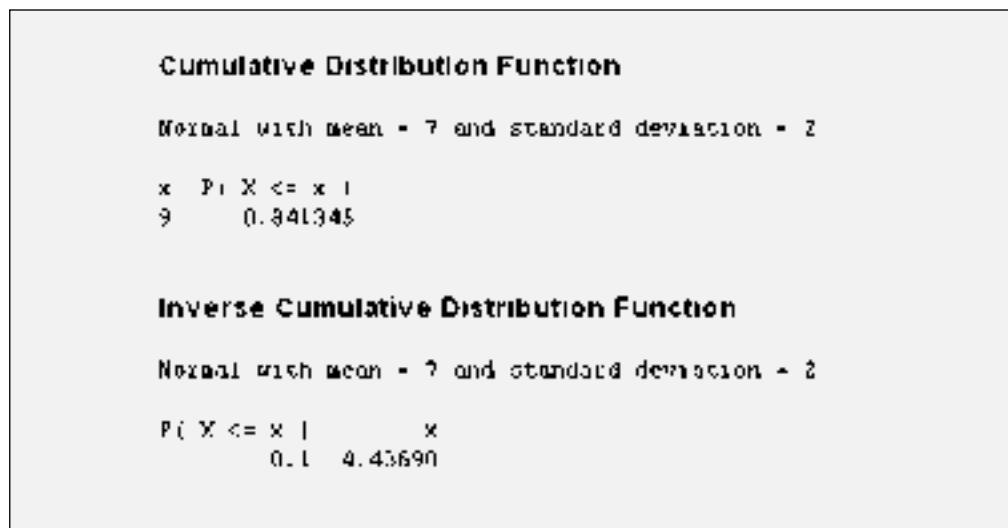
Hoja de trabajo de Excel para calcular probabilidades normales.

A	B
1 Normal Probabilities	
2	
3 Common Data	
4 Mean	7
5 Standard Deviation	2
6	
7 Probability for $X \leq$	
8 X Value	3.5
9 Z Value	-1.75
10 $P(X \leq 3.5)$	0.9401
11	
12 Find X and Z Given Cum. Pctage.	
13 Cumulative Percentage	10.00%
14 Z Value	-1.2816
15 X Value	4.4369

=STANDARDIZE(B8, B4, B5)  
 =NORMDIST(B8, B4, B5, TRUE)  
 =NORMSINV(B13)  
 =NORMINV(B13, B4, B5)

**FIGURA 6.19**

Probabilidades normales en Minitab.



## PROBLEMAS PARA LA SECCIÓN 6.2

### Aprendizaje básico

**ASISTENCIA** de PH Grade **6.1** Dada una distribución normal (con una media de 0 y una desviación estándar de 1, como en la tabla E.2), ¿cuál es la probabilidad de que

- Z sea menor que 1.57?
- Z sea mayor que 1.84?
- Z esté entre 1.57 y 1.84?
- Z sea menor que 1.57 o mayor que 1.84?

**6.2** Dada una distribución normal estandarizada (con una media de 0 y una desviación estándar de 1, como en la tabla E.2), ¿cuál es la probabilidad de que

- Z esté entre -1.57 y 1.84?
- Z sea menor que -1.57 o mayor a 1.84?
- ¿Cuál es el valor de Z si sólo el 2.5% de todos los posibles valores de Z son más grandes?
- Entre cuáles dos valores de Z (distribuidos simétricamente alrededor de la media) estarán contenidos el 68.26% de todos los posibles valores Z?

**ASISTENCIA** de PH Grade

**6.3** Dada una distribución normal estandarizada (con una media de 0 y una desviación estándar de 1 como en la tabla E.2), ¿cuál es la probabilidad de que

- Z sea menor que 1.08?
- Z sea mayor que -0.21?
- Z sea menor que -0.21 o mayor que la media?
- Z sea menor que -0.21 o mayor que 1.08?

**6.4** Dada una distribución normal estandarizada (con una media de 0 y una desviación estándar de 1 como en la tabla E.2), determine las siguientes probabilidades:

- $P(Z > 1.08)$
- $P(Z < -0.21)$
- $P(-1.96 < Z < -0.21)$
- ¿Cuál es el valor de Z si sólo 15.87% de todos los valores posibles de Z son mayores?

**ASISTENCIA** de PH Grade

**6.5** Dada una distribución normal con  $\mu = 100$  y  $\sigma = 10$ , ¿cuál es la probabilidad de que

- $X > 75$ ?

- b.  $X < 70$ ?
- c.  $X < 80$  o  $X > 110$ ?
- d. el 80% de los valores estén entre los dos valores  $X$  (simétricamente distribuidos alrededor de la media)?

**ASISTENCIA** de PH Grade **6.6** Dada una distribución normal con  $\mu = 50$  y  $\sigma = 4$ , ¿cuál es la probabilidad de que

- a.  $X > 43$ ?
- b.  $X < 42$ ?
- c. el 5% de los valores sean menores que el valor  $X$ ?
- d. el 60% de los valores estén entre los dos valores  $X$  (simétricamente distribuidos alrededor de la media)?

### Aplicación de conceptos

**6.7** Durante 2001, el 61.3% de los hogares en Estados Unidos compraron café de grano y gastaron un promedio de \$36.16 en café de grano durante el año (“Annual Product Preference Study”, *Progressive Grocer*, 1 de mayo, 2002, 31). Considere el desembolso anual para café de grano en los hogares que lo compran, suponiendo que este desembolso se distribuye aproximadamente como una variable aleatoria normal con una media de \$36.16 y una desviación estándar de \$10.00.

- a. Encuentre la probabilidad de que un hogar gaste menos de \$25.00.
- b. Encuentre la probabilidad de que un hogar gaste más de \$50.00.
- c. ¿Qué proporción de los hogares gastan entre \$30.00 y \$40.00?
- d. ¿El 99% de los hogares gastan menos de qué cantidad?

**ASISTENCIA** de PH Grade **AUTO** Examen **6.8** La empresa Toby’s Truck determina que, en una base anual, la distancia recorrida por camión se distribuye normalmente con una media de 50.0 mil millas y una desviación estándar de 12.0 mil millas.

- a. ¿Qué proporción de camiones se espera que recorran entre 34.0 y 50.0 mil millas en el año?
- b. ¿Qué porcentaje de camiones se espera que recorran por debajo de 30.0 o por arriba de 60.0 mil millas en el año?
- c. ¿Cuántas millas habrán sido recorridas por al menos el 80% de los camiones?
- d. ¿Cuáles serían sus respuestas a los incisos a) a c) si la desviación estándar fuera de 10.0 mil millas?

**ASISTENCIA** de PH Grade **6.9** La fuerza de rompimiento de las bolsas de plástico usadas para empacar productos se distribuye normalmente con una media de 5 libras por pulgada cuadrada y una desviación estándar de 1.5 libras por pulgada cuadrada. ¿Qué proporción de bolsas tienen una fuerza de rompimiento de

- a. menos de 3.17 libras por pulgada cuadrada?
- b. al menos 3.6 libras por pulgada cuadrada?
- c. entre 5 y 5.5 libras por pulgada cuadrada?
- d. ¿Entre cuáles dos valores simétricamente distribuidos alrededor de la media se encontrarán el 95% de las fuerzas de rompimiento?

**6.10** Un conjunto de calificaciones finales para un curso de Introducción a la estadística se distribuye normalmente con una media de 73 y una desviación estándar de 8.

- a. ¿Cuál es la probabilidad de que un alumno obtenga una calificación de 91 o menos en este examen?
- b. ¿Cuál es la probabilidad de que un alumno obtenga una calificación entre 65 y 89?
- c. Hay una probabilidad del 5% de que un alumno que realice el examen obtenga notas mayores a qué calificación?
- d. Si el profesor califica con base en una curva (da 10 al 10% más alto de la clase sin importar la calificación), ¿un alumno estará mejor con una calificación de 81 en este examen o con una calificación de 68 en un examen diferente donde la media es de 62 y la desviación estándar es de 3? Muestre estadísticamente su respuesta y explique.

**6.11** Un análisis estadístico de 1,000 llamadas de larga distancia realizadas desde las oficinas de la corporación Bricks and Clicks Computer Corporation indica que la duración de estas llamadas se distribuye normalmente con una media de  $\mu = 240$  segundos y  $\sigma = 40$  segundos.

- a. ¿Cuál es la probabilidad de que una llamada haya durado menos de 180 segundos?
- b. ¿Cuál es la probabilidad de que una llamada específica haya durado entre 180 y 300 segundos?
- c. ¿Cuál es la probabilidad de que una llamada haya durado entre 110 y 180 segundos?
- d. ¿Cuál es la duración de una llamada en particular si sólo el 1% de todas las llamadas son más cortas?

**6.12** Al número de acciones negociadas diariamente en la Bolsa de Valores de Nueva (NYSE) se le conoce como el *volumen* negociado. El 23 de abril de 2004 se negociaron 1.395 miles de millones de acciones (“NYSE Volume”, *The Wall Street Journal*, 26 de abril, 2004, C2). Este volumen de negociaciones se acerca a la media del volumen para la NYSE. Suponga que el número de acciones negociadas en la NYSE es una variable aleatoria normal con una media de 1.4 miles de millones y una desviación estándar de 0.15 miles de millones. Para un día seleccionado aleatoriamente, ¿cuál es la probabilidad de que el volumen negociado de la NYSE sea:

- a. menor a 1.7 miles de millones?
- b. menor a 1.25 miles de millones?
- c. menor a 1.0 mil millones?
- d. mayor a 1.0 mil millones?

**6.13** Muchos problemas de producción se relacionan con la unión exacta de partes de maquinaria, como flechas, que caben en el orificio de una válvula. Un diseño en particular requiere de una flecha con un diámetro de 22.00 mm, pero las flechas con diámetros entre 21.900 mm, y 22.010 mm son aceptables. Suponga que el proceso de manufactura fabrica flechas con diámetros que se distribuyen normalmente con una media de 22.002 mm y con una desviación estándar de 0.005 mm. Para este proceso, ¿cuál es

- a. la proporción de flechas con un diámetro entre 21.90 mm y 22.00 mm?
- b. la probabilidad de que una flecha sea aceptada?
- c. el diámetro que será sólo el 2% de las flechas excederán?
- d. ¿Cuáles serían sus respuestas en los incisos a) a c) si la desviación estándar de los diámetros de las flechas fuera de 0.004 mm?

## 6.3 EVALUACIÓN DE LA NORMALIDAD

Como se discutió en la sección 6.2, muchas de las variables continuas usadas en negocios se asemejan estrechamente a una distribución normal. Sin embargo, muchas variables importantes ni siquiera pueden ser aproximadas por la distribución normal. Esta sección presenta dos modelos para evaluar si un conjunto de datos puede ser aproximado por una distribución normal:

1. Compare las características del conjunto de datos con las propiedades de la distribución normal.
2. Realice un plano de probabilidad normal.

### Evaluación de las propiedades

La distribución normal tiene varias propiedades teóricas importantes:

- Es simétrica, por lo tanto, la media y la mediana son iguales.
- Tiene forma de campana, por lo que se aplica la regla empírica.
- El rango intercuartil es igual a 1.33 desviaciones estándar.
- El rango es infinito.

En la práctica, algunas variables continuas tienen características que se acercan a las propiedades teóricas. Sin embargo, muchas variables continuas no son distribuidas normalmente, ni tampoco distribuidas aproximadamente. Para tales variables, las características descriptivas de los datos no corresponden bien con las propiedades de la distribución normal. Un enfoque para verificar la normalidad consiste en comparar las características de los datos actuales con las propiedades correspondientes que subyacen a la distribución normal, como sigue.

- Construya gráficas y observe su apariencia. Para conjuntos de datos pequeños o de tamaño moderado, diseñe un diagrama de tallo y hojas o una gráfica de caja y bigote. Para conjuntos de datos más grandes, elabore la distribución de frecuencia y trace el histograma o polígono.
- Calcule medidas numéricas descriptivas y compare las características de los datos con las propiedades teóricas de una distribución normal. Compare la media y la mediana. ¿El rango intercuartil es aproximadamente 1.33 veces la desviación estándar? ¿Es el rango aproximadamente 6 veces la desviación estándar?
- Evalúe cómo se distribuyen los datos. Determine si aproximadamente dos tercios de los valores caen entre la media  $\pm 1$  desviación estándar. Determine si aproximadamente cuatro quintos de los valores caen entre la media  $\pm 1.28$  desviaciones estándar. Determine si aproximadamente 19 de cada 20 valores caen entre la media  $\pm 2$  desviaciones estándar.

¿Los rendimientos en 2003 de los fondos de inversión que se analizaron en los capítulos 2 y 3 tienen las propiedades de la distribución normal? La figura 6.20 muestra la estadística descriptiva para estos datos y la figura 6.21 presenta una gráfica de caja y bigote.

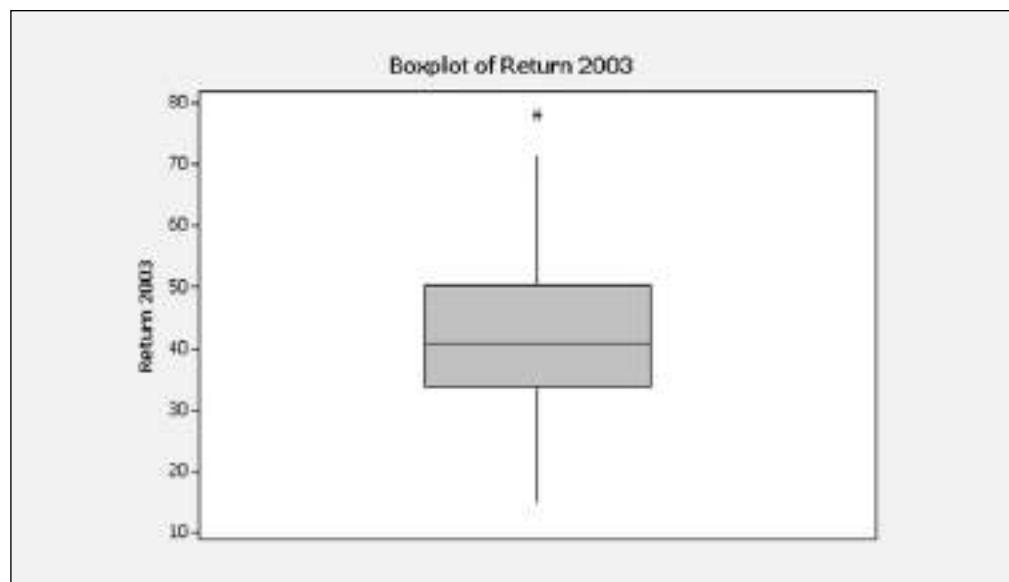
**FIGURA 6.20**

Estadística descriptiva en Excel para los rendimientos en 2003 de los fondos de inversión.

A	B
1	<i>Return 2003</i>
2	
3	<b>Mean</b> 42.6215
4	<b>Standard Error</b> 1.1852
5	<b>Median</b> 40.8
6	<b>Mode</b> 37.5
7	<b>Standard Deviation</b> 13.0369
8	<b>Sample Variance</b> 169.9609
9	<b>Kurtosis</b> -0.2828
10	<b>Skewness</b> 0.3500
11	<b>Range</b> 63.1
12	<b>Minimum</b> 14.9
13	<b>Maximum</b> 78
14	<b>Sum</b> 5157.2
15	<b>Count</b> 121
16	<b>Largest(1)</b> 78
17	<b>Smallest(1)</b> 14.9

**FIGURA 6.21**

Gráfica de caja y bigote en Minitab para los rendimientos 2003 de los fondos de inversión.



De estas figuras se desprenden las siguientes afirmaciones:

1. La media de 42.62 es ligeramente mayor a la mediana de 40.8.
2. La gráfica de caja y bigote aparece ligeramente sesgada hacia la derecha con un valor extremo, el 78.
3. El rango intercuartilar de 16.45 está aproximadamente a 1.33 desviaciones estándar.
4. El rango de 63.1 es igual a 4.84 desviaciones estándar.
5. El 65.6% de los rendimientos están dentro de  $\pm 1$  desviación estándar de la media.
6. El 78.7% de los rendimientos están dentro de  $\pm 1.28$  desviaciones estándar de la media.

Con base en estas afirmaciones y en los criterios anteriormente dados, se concluye que los rendimientos para 2003 están aproximadamente distribuidos de forma normal. Sin embargo, las afirmaciones 1 y 2 indican que los rendimientos en 2003 están sesgados ligeramente hacia la derecha.

### Construcción de un plano de probabilidad normal

Un **plano de probabilidad normal** es un acercamiento gráfico para evaluar si los datos están distribuidos normalmente. Un acercamiento común es la **gráfica cuantil-cuantil**. En este método, cada valor ordenado se transforma en una puntuación  $Z$ , y después se grafican los valores de los datos contra las puntuaciones  $Z$ . Por ejemplo, si se tiene una muestra de  $n = 19$ , el valor  $Z$  para el menor valor corresponde al área acumulativa de  $\frac{1}{n+1} = \frac{1}{19+1} = \frac{1}{20} = 0.05$ . El valor  $Z$  para el área acumulativa de 0.05 (de la tabla E.2) es  $-1.65$ . La tabla 6.7 ilustra el conjunto completo de valores  $Z$  para una muestra de  $n = 19$ .

**TABLA 6.7**

Valores ordenados y sus valores  $Z$  correspondientes para una muestra de  $n = 19$ .

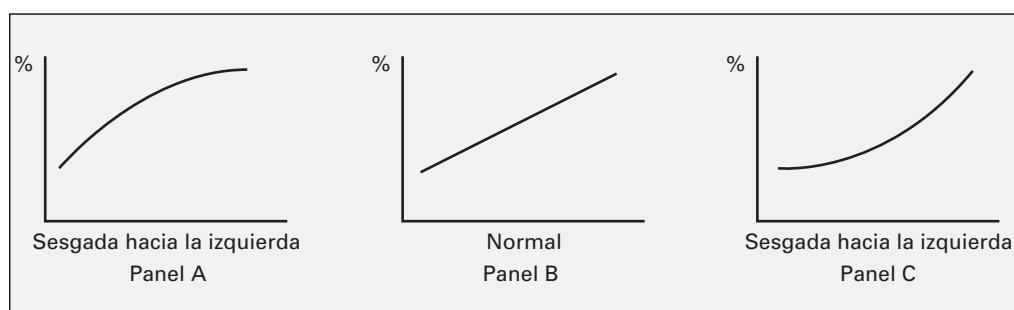
Valor ordenado	Valor $Z$	Valor ordenado	Valor $Z$
1	-1.65	11	0.13
2	-1.28	12	0.25
3	-1.04	13	0.39
4	-0.84	14	0.52
5	-0.67	15	0.67
6	-0.52	16	0.84
7	-0.39	17	1.04
8	-0.25	18	1.28
9	-0.13	19	1.65
10	0.00		

Los valores  $Z$  se trazan en el eje  $X$  y los valores correspondientes de la variable se trazan en el eje  $Y$ . Si los datos se distribuyen de forma normal, los puntos se trazarán aproximadamente a lo largo de una línea recta.

Un segundo modelo (usado por Minitab) transforma el eje vertical  $Y$  de una manera un poco más complicada que va más allá del objetivo de este texto. Una vez más, si los datos se distribuyen de forma normal, los puntos se trazarán aproximadamente a lo largo de una línea recta. La figura 6.22 ilustra la forma típica de las gráficas de probabilidad normal para una distribución sesgada hacia la izquierda (panel A), una distribución normal (panel B), y una distribución sesgada hacia la derecha (panel C). Si los datos están sesgados hacia la izquierda, la curva se elevará más rápidamente al inicio y después disminuirá. Si los datos están sesgados hacia la derecha, los datos se elevarán lentamente al inicio y después se elevarán a una tasa más rápida para los valores más altos de la variable a trazar.

**FIGURA 6.22**

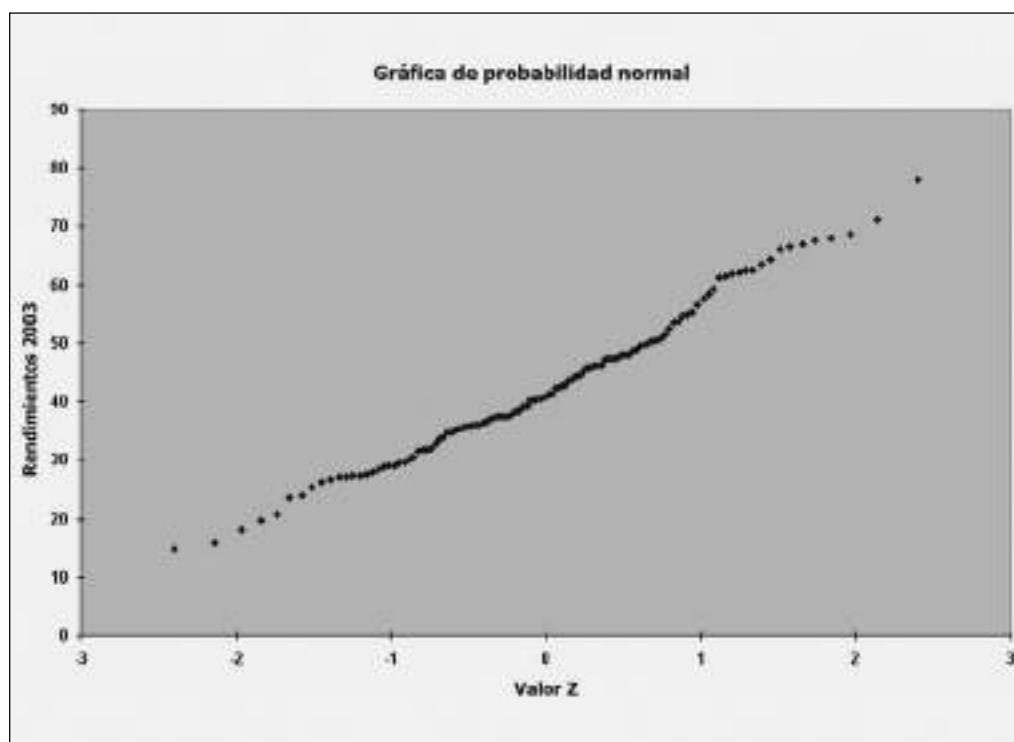
Gráficas de probabilidad normal para una distribución sesgada hacia la izquierda, de una distribución normal y de una distribución sesgada hacia la derecha.



La figura 6.23 muestra una gráfica de probabilidad normal cuantil-cuantil de Excel y la figura 6.24 muestra una gráfica de probabilidad normal de Minitab para los rendimientos en 2003.

**FIGURA 6.23**

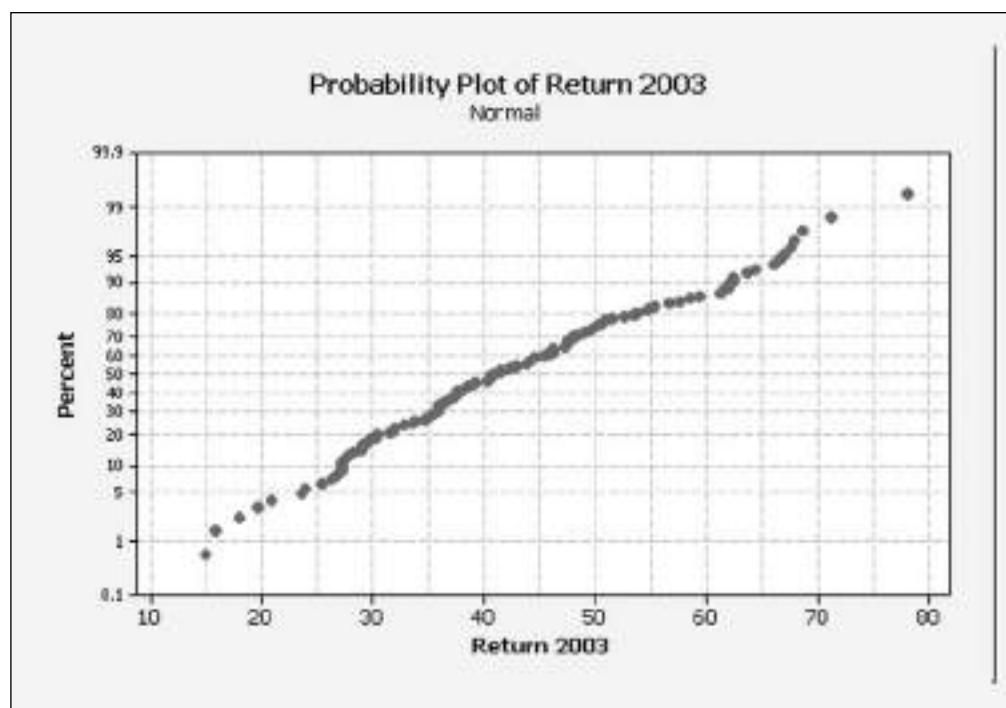
Gráfica de probabilidad normal de Excel para los rendimientos en 2003.



Las figuras 6.23 y 6.24 muestran que la gráfica de probabilidad normal de los rendimientos en 2003 se aproxima a una línea recta. Se concluye que los rendimientos en 2003 están distribuidos aproximadamente de forma normal.

**FIGURA 6.24**

Gráfica de probabilidad normal de Minitab para los rendimientos en 2003.



## PROBLEMAS PARA LA SECCIÓN 6.3

### Aprendizaje básico

**ASISTENCIA de PH Grade** **6.14** Pruebe que para una muestra de  $n = 39$ , los valores Z más pequeños y más grandes son  $-1.96$  y  $+1.96$ , y el valor Z medio (es decir,  $20^\circ$ ) es  $0.00$ .

**6.15** Para una muestra de  $n = 6$ , liste los seis valores Z.

### Aplicación de conceptos

Puede resolver los problemas 6.16 a 6.19 manualmente o en Excel, Minitab o SPSS. Le recomendamos resolver los problemas 6.20 a 6.22 con Excel, Minitab o SPSS.

**AUTO Examen** **6.16** El precio diario de hotel para 25 ciudades en marzo de 2004 HOTEL-PRICE fue el siguiente:

Ciudad	Precio de hotel
Anaheim	95.26
Atlanta	78.91
Boston	112.92
Chicago	96.90
Dallas	77.43
Denver	74.22
Detroit	77.71
Filadelfia	95.02
Houston	76.26
Los Ángeles	95.78
Miami	140.61
Minneapolis	78.64
Nashville	74.61

Ciudad	Precio de hotel
Nueva Orleáns	121.59
Nueva York	167.43
Norfolk	62.88
Oahu Island	119.76
Orlando	98.57
Phoenix	123.19
San Diego	110.23
San Francisco	123.51
Seattle	95.09
St. Louis	74.68
Tampa	97.08
Washington	123.27

Fuente: USA Today, 27 de abril, 2004, 5B.

Determine si los datos parecen o no distribuirse aproximadamente de forma normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- construyendo una gráfica de probabilidad normal.

**6.17** Un problema en una línea telefónica que evita que el cliente reciba o realice llamadas es desconcertante tanto para el cliente como para la empresa telefónica. Los datos en la parte superior de la página 198 representan dos muestras de 20 problemas reportados a dos diferentes oficinas de la empresa de teléfonos. El tiempo que se tardan en reparar estos problemas en las líneas telefónicas de los clientes se registra en minutos. PHONE

Oficina central I. Tiempo para reparar problemas (minutos)										
1.48	1.75	0.78	2.85	0.52	1.60	4.15	3.97	1.48	3.10	
1.02	0.53	0.93	1.60	0.80	1.05	6.32	3.93	5.45	0.97	
Oficina central II. Tiempo para reparar problemas (minutos)										
7.55	3.75	0.10	1.10	0.60	0.52	3.30	2.10	0.58	4.02	
3.75	0.65	1.92	0.60	1.53	4.23	0.08	1.48	1.65	0.72	

Determine, para cada una de las dos locaciones, si los datos parecen distribuirse de forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- realizando una gráfica de probabilidad normal.

**6.18** Muchos procesos de manufactura usan el término trabajo en proceso (con frecuencia abreviado WIP, por las siglas en inglés para “work-in-process”). En una compañía impresora de libros, el WIP representa el tiempo necesario para doblar, juntar, coser, pegar por un extremo y encuadrinar las hojas procedentes de la prensa. Los siguientes datos representan muestras de 20 libros en cada una de las dos plantas de producción y el tiempo de procesamiento (definido operacionalmente como el tiempo en días desde que los libros salieron de la prensa hasta que fueron empacados en cajas de cartón) para este trabajo: **WIP**

Planta A										
15.62	5.29	16.25	10.92	11.46	21.62	8.45	8.58	5.41	11.42	
11.62	7.29	17.50	17.96	14.42	10.50	7.58	9.29	7.54	18.92	
Planta B										
9.54	11.46	16.62	12.62	25.75	15.41	14.29	13.13	13.71	10.04	
5.75	12.46	19.17	13.21	16.00	12.33	14.25	15.37	16.25	19.71	

Determine, para cada una de las dos plantas, si los datos parecen o no distribuirse en forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- construyendo una gráfica de probabilidad normal.

**6.19** Las puntuaciones de crédito son números de tres dígitos usados por los prestamistas para evaluar nuestro valor de crédito. Las puntuaciones para los residentes de veinte áreas metropolitanas son las siguientes: **CREDITSCORE**

Ciudad	Puntuación de crédito
Atlanta	670
Boston	705
Chicago	680
Cleveland	690
Dallas	653
Denver	675
Detroit	675
Houston	655
Los Ángeles	667
Miami	672
Minneapolis	707

Ciudad	Puntuación de crédito
Nueva York	688
Orlando	671
Filadelfia	688
Phoenix	660
Sacramento	676
San Francisco	686
Seattle	691
Tampa	675
Washington	693

Determine si los datos parecen o no distribuirse de forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- construyendo una gráfica de probabilidad normal.

**6.20** Una de las operaciones de un molino consiste en cortar trozos de acero en partes que posteriormente serán usadas como marco para los asientos delanteros en una planta automotriz. El acero se corta con una sierra de diamante y se requiere que las partes resultantes sean más o menos de 0.005 pulgadas de la longitud especificada por la empresa automotriz. Los datos provienen de una muestra de 100 partes de acero. **STEEL** La medida que se reporta es la diferencia en pulgadas entre la longitud real de la parte de acero, medida con láser, y la longitud especificada de la parte de acero.

Determine si los datos parecen o no distribuirse de forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- realizando una gráfica de probabilidad normal.

**6.21** En una fábrica que produce esquinas de goma, la goma cruda se mezcla en una máquina moldeadora y después se corta en tiras. Las tiras se cargan en una máquina de moldes y fundidas en las formas deseadas de esquinas de goma. Los pesos (en gramos) de la muestra de esquinas de goma son los siguientes: **RUBBER**

8.63	8.59	8.63	8.67	8.64	8.57	8.53	8.59	8.66	8.54
8.65	8.61	8.67	8.65	8.64	8.64	8.51	8.61	8.65	8.62
8.57	8.60	8.54	8.69	8.52	8.63	8.72	8.58	8.66	8.66
8.57	8.66	8.62	8.66	8.69	8.57	8.58	8.65	8.68	8.56
8.54	8.65	8.65	8.62	8.66	8.61	8.64	8.73	8.62	8.60
8.69	8.50	8.58	8.63	8.66	8.59	8.69	8.70	8.54	8.62
8.63	8.61	8.65	8.59	8.61	8.56	8.64	8.65	8.67	8.61
8.64	8.61	8.67	8.65	8.55	8.71	8.75	8.56	8.62	8.66

Fuente: W. L. Pearn and K. S. Chen, “A Practical Implementation of the Process Capability Index Cpk”, Quality Engineering, 1997, 9, 721-737.

Determine si los datos parecen o no distribuirse de forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- construyendo una gráfica de probabilidad normal.

**6.22** Los siguientes datos representan el costo de la electricidad en dólares durante el mes de julio de 2004 para una muestra aleatoria de 50 departamentos de dos habitaciones en una ciudad grande: **UTILITY**

96	171	202	178	147	102	153	197	127	82
157	185	90	116	172	111	148	213	130	165
141	149	206	175	123	128	144	168	109	167
95	163	150	154	130	143	187	166	139	149
108	119	183	151	114	135	191	137	129	158

Determine si los datos parecen o no distribuirse de forma aproximadamente normal:

- evaluando las propiedades reales frente a las propiedades teóricas.
- construyendo una gráfica de probabilidad normal.

## RESUMEN

En este capítulo usted utilizó la distribución normal en el escenario “Uso de la estadística” para estudiar el tiempo de descarga de una página Web. Además, estudió la gráfica de probabili-

dad normal. En el siguiente capítulo, usará la distribución normal para desarrollar el tema de la estadística inferencial.

## FÓRMULAS IMPORTANTES

### La función de densidad de la probabilidad normal

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(X-\mu)/\sigma]^2} \quad (6.1)$$

### Encontrar el valor Z (fórmula de transformación)

$$Z = \frac{X - \mu}{\sigma} \quad (6.2)$$

### La función de densidad de probabilidad normal estandarizada

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)Z^2} \quad (6.3)$$

### Encontrar el valor X

$$X = \mu + Z\sigma \quad (6.4)$$

## CONCEPTOS CLAVE

distribución normal 179

distribución normal estandarizada

acumulativa 181

fórmula de transformación 181

función de densidad de probabilidad continua 178

función de densidad de probabilidad normal 180

gráfica cuantil-cuantil 195

gráfica de probabilidad normal 195

variable aleatoria normal

estandarizada Z 181

## PROBLEMAS DE REPASO

### Revisión de su comprensión

**6.23** ¿Por qué es que sólo una tabla de distribución normal como la tabla E.2 es necesaria para encontrar cualquier probabilidad por debajo de la curva normal?

**6.24** ¿Cómo se encuentra el área entre dos valores debajo de la curva normal?

**6.25** ¿Cómo se encuentra el valor  $X$  que corresponde a un percentil dada una distribución normal?

**6.26** ¿Cómo se utiliza la gráfica de probabilidad normal para evaluar si un conjunto de datos está distribuido normalmente?

### Aplicación de conceptos

**6.27** Una máquina de coser industrial usa cojinetes que deben tener un diámetro de 0.75 pulgadas. Los límites inferior y superior especificados dentro de los que los cojinetes pueden operar son 0.74 y 0.76 pulgadas, respectivamente. La experiencia indica que el diámetro real de los cojinetes se distribuye de forma aproximadamente normal con una media de 0.753 pulgadas y una desviación estándar de 0.004 pulgadas. ¿Cuál es la probabilidad de que el cojinete esté

- entre el diámetro objetivo y la media real?
- entre el límite inferior especificado y el diámetro objetivo?
- por encima del límite superior especificado?

- d. por debajo del límite inferior especificado?
- e. el 93% de los diámetros son mayores, a qué valor?

**6.28** La cantidad de líquido contenido en botellas de bebida refrescante se distribuye normalmente con una media de 2.0 litros y una desviación estándar de 0.05 litros. Las botellas que contienen menos del 95% del contenido neto listado (1.90 litros en este caso) son causa de que los productores sean penalizados por la oficina estatal de asuntos del consumidor. Las botellas que tienen un contenido neto superior a 2.10 litros, pueden causar un exceso de derrame cuando se abren. ¿Cuál es la proporción de botellas que contendrán:

- a. entre 1.90 y 2.0 litros?
- b. entre 1.90 y 2.10 litros?
- c. menos de 1.90 o más de 2.10 litros?
- d. el 99% de las botellas contienen por lo menos ¿qué cantidad de refresco?
- e. el 99% de las botellas contendrán una cantidad que está ¿entre cuáles dos valores (simétricamente distribuidos) alrededor de la media?

**6.29** En un esfuerzo por reducir el número de botellas que contienen menos de 1.90 litros, el embotellador del problema 6.28 programa la máquina que llena los envases para que la media sea de 2.02 litros. En estas circunstancias, ¿cuáles serían sus respuestas para los incisos a) al e)?

**6.30** Un productor de jugo de naranja compra todas sus naranjas de un gran naranjal. La cantidad de jugo exprimido de cada una de estas naranjas se distribuye aproximadamente de forma normal con una media de 4.70 onzas y una desviación estándar de 0.40 onzas.

- a. ¿Cuál es la probabilidad de que una naranja elegida aleatoriamente contenga entre 4.70 y 5.00 onzas?
- b. ¿Cuál es la probabilidad de que una naranja elegida aleatoriamente contenga entre 5.00 y 5.50 onzas?
- c. El 77% de las naranjas contendrán por lo menos ¿cuántas onzas de jugo?
- d. ¿Entre cuáles dos valores (en onzas) simétricamente distribuidos alrededor de la media poblacional, estarán el 80% de las naranjas?

**6.31** De acuerdo con *Investment Digest* ("Diversification and the Risk/Reward Relationship", Invierno de 1994, 1-3), la media del rendimiento anual para las acciones comunes de 1926 a 1992 fue del 12.4% y la desviación estándar del rendimiento anual fue del 20.6%. El artículo asegura que la distribución de los rendimientos anuales para las acciones comunes es aproximadamente simétrica y con forma de campana. Suponga que la distribución es de forma normal con la media y desviación estándar indicadas arriba. Encuentre la probabilidad de que los rendimientos para las acciones comunes sea

- a. mayor del 0%.
- b. mayor del 10%.
- c. mayor del 20%.
- d. menor del -10%.

**6.32** Durante este mismo lapso de 67 años mencionado en el problema 6.31, la media del rendimiento anual para los bonos del gobierno a largo plazo fue de 5.2, y la desviación estándar fue de 8.6%. El artículo afirma que la distribución de los rendimientos anuales de los bonos de gobierno a largo plazo es aproximadamente simétrica y con forma de campana. Suponga que

la distribución se distribuye normalmente con la media y desviación estándar indicadas arriba. Encuentre la probabilidad de que el rendimiento para los bonos de gobierno a largo plazo será

- a. mayor del 0%.
- b. mayor del 10%.
- c. mayor del 20%.
- d. menor del -10%.
- e. Discuta las diferencias entre los rendimientos anuales de las acciones comunes y los bonos de gobierno a largo plazo.

**6.33** El *Wall Street Journal* reportó que casi todos los principales índices de acciones habían tenido fuertes ganancias en los últimos 12 meses ("What's Hot... and Not", *The Wall Street Journal*, 26 de abril, 2004, C3). El rendimiento anual del S&P 500, que comprende 500 grandes empresas, fue de aproximadamente del +27%. Los rendimientos de un año de Russell 2000, que agrupa a 2000 pequeñas empresas, fue de aproximadamente el +52%. A lo largo de la historia, los rendimientos de un año son aproximadamente normales. La desviación estándar para los rendimientos S&P 500 es aproximadamente del 20% y para Russell 2000 la desviación estándar es aproximadamente del 35%.

- a. ¿Cuál es la probabilidad de que una acción de S&P 500 haya ganado el 30% o más en el último año? ¿Que haya ganado el 60% o más en el último año?
- b. ¿Cuál es la probabilidad de que una acción de S&P 500 perdiera dinero el año pasado? ¿Y de que perdiera el 30% o más dinero?
- c. Repita los incisos a) y b) para las acciones en Russell 2000.
- d. Escriba un breve resumen acerca de lo que haya encontrado. Asegúrese de incluir una explicación acerca de los riesgos asociados a tener una desviación estándar grande.

**6.34** El *New York Times* reportó (Laurie J. Flynn, "Tax Surfing", *The New York Times*, 25 de marzo, 2002, C10) que la media del tiempo de descarga para la página principal del sitio Web Internal Revenue Service [www.irs.gov](http://www.irs.gov) es de 0.8 segundos. Suponga que el tiempo de descarga se distribuye normalmente con una desviación estándar de 0.2 segundos. ¿Cuál es la probabilidad de que el tiempo de descarga sea

- a. menor a 1 segundo?
- b. de entre 0.5 y 1.5 segundos?
- c. de más de 0.5 segundos?
- d. el 99% de los tiempos de descarga están por arriba ¿de cuántos segundos?
- e. ¿Entre cuáles dos valores simétricamente distribuidos alrededor de la media se encuentra el 95% de los tiempos de descarga?

**6.35** El mismo artículo mencionado en el problema 6.34 también reportó que la media del tiempo de descarga para el sitio Web de H&R Block [www.hrblock.com](http://www.hrblock.com) es de 2.5 segundos. Suponga que el tiempo de descarga se distribuye normalmente con una desviación estándar de 0.5 segundos. ¿Cuál es la probabilidad de que el tiempo de descarga sea

- a. menor a 1 segundo?
- b. de entre 0.5 y 1.5 segundos.
- c. de más de 0.5 segundos?
- d. ¿Por arriba de cuántos segundos se encuentran el 99% de los tiempos de descarga?
- e. Compare los resultados para el sitio IRS calculados en el problema 6.34 con los del sitio H&R Block.

**6.36 (Proyecto de grupo)** De acuerdo con Burton G. Malkiel, los cambios diarios en el precio de cierre de las acciones sigue un *camino aleatorio* (es decir, estos eventos diarios son independientes uno del otro y se mueven hacia arriba o hacia debajo de forma aleatoria) y pueden aproximarse a una curva normal. Para probar esta teoría, utilice el periódico o Internet para seleccionar una empresa que cotice en la Bolsa de Valores de Nueva York, una empresa que cotice en el American Stock Exchange, y una empresa que cotice “sobre el escritorio” (es decir, en el mercado nacional NASDAQ) y después haga lo siguiente:

1. Registre los cambios diarios del precio de cierre de las acciones para cada una de las empresas durante 6 semanas consecutivas (para que así tenga 30 valores por empresa).
  2. Registre los cambios diarios del cierre del precio de las acciones para cada una de estas empresas durante 6 semanas consecutivas (para que así tenga 30 valores por empresa).
- Para cada uno de sus seis conjuntos de datos, determine si los datos se distribuyen de forma aproximadamente normal
- a. examinando una gráfica de tallo y hoja, un histograma o polígono y una gráfica de caja y bigote.
  - b. evaluando las propiedades reales frente a las propiedades teóricas.
  - c. realizando una gráfica de probabilidad normal.
  - d. Analice los resultados de los incisos a), b) y c). ¿Qué concluye ahora sobre sus tres compañías con respecto a los precios de cierre diarios y los cambios en los precios de cierre? ¿Cuál de los conjuntos de datos, si acaso alguno, se aproxima a una distribución normal?

*Nota:* La teoría del *camino aleatorio* pertenece a los *cambios* diarios en el precio de cierre de las acciones, no al precio diario de cierre de las acciones.



## PROYECTO EN EQUIPO

Los datos del archivo **MUTUALFUNDS2004** contienen información respecto a 12 variables de una muestra de 121 fondos de inversión. Las variables son:

- Fund —El nombre del fondo de inversión.
- Category —Tipo de acciones comprendidas en el fondo de inversión: pequeño capital, capital medio, gran capital.
- Objective —Objetivo de las acciones comprendidas en el fondo de inversión: crecimiento o valor.
- Assets —Activos en millones de dólares.
- Fees —Cargos por ventas (sí o no).
- Expense ratio —Relación entre gastos y activos netos en porcentaje.
- 2003 Return —Rendimiento de 12 meses en el 2003.
- Three-year return —Rendimiento anualizado de 2001 a 2003.
- Five-year return —Rendimiento anualizado de 1999 a 2003.
- Risk —Factor riesgo-pérdida de los fondos de inversión clasificado como bajo, promedio o alto.
- Best quarter —Mejor rendimiento trimestral de 1999 a 2003.
- Worst quarter —Peor rendimiento trimestral de 1999 a 2003.

**6.37** Considere las variables de relación de gastos, rendimiento anualizado a tres años y rendimiento anualizado a cinco años. Para cada una de estas variables, determine si los datos se distribuyen aproximadamente de forma normal

- a. evaluando las propiedades reales frente a las teóricas.
- b. construyendo una gráfica de probabilidad normal.

## CASO ACTUAL ADMINISTRACIÓN DEL SPRINGVILLE HERALD

El departamento de producción del periódico se ha embarcado en un esfuerzo por mejorar la calidad. Su primer proyecto se relaciona con la tonalidad oscura de la impresión del periódico. Cada día se necesita determinar qué tan oscura es la impresión. La tonalidad se mide en una escala estándar en la que el valor objetivo es 1.0. Los datos recopilados en el último año indican que la tonalidad negra se distribuye normalmente con una media de 1.005 y una desviación estándar de 0.10.

Cada día se elige una mancha del primer periódico impreso y se mide la tonalidad oscura. Ésta se considera aceptable si está entre 0.95 y 1.05.

### EJERCICIO

SH6.1 Suponiendo que la distribución no ha cambiado con respecto a la del año pasado, ¿cuál es la probabilidad de que la tonalidad de la mancha sea:

- a. menor a 1.0?
- b. entre 0.95 y 1.0?
- c. entre 1.0 y 1.05?
- d. menos de 0.95 o mayor que 1.05?

SH6.2 El objetivo del equipo de producción es reducir la probabilidad de que la tonalidad oscura esté por debajo de 0.95 o por arriba de 1.05. ¿Sería mejor enfocarse en el proceso de mejoramiento que disminuyó la media del valor objetivo de 1.0 o en el proceso de mejoramiento que redujo la desviación estándar a 0.075? Explique por qué.

## CASO WEB

Aplique su conocimiento acerca de la distribución normal en este caso Web que viene del escenario “Uso de la estadística” de este capítulo.

Para satisfacer las preocupaciones de los anunciantes potenciales, la administración de OnCampus! ha puesto en marcha un proyecto de investigación para aprender sobre la cantidad de tiempo que los lectores permanecen en sus sitios Web. El departamento de marketing ha recopilado datos y ha hecho algunas aseveraciones basadas en la afirmación de que los datos siguen una distribución normal. Los datos y las conclusiones pueden encontrarse en un reporte dentro del si-

tio Web interno [www.prenhall.com/Springville/OC\\_MarketingSurvey.htm](http://www.prenhall.com/Springville/OC_MarketingSurvey.htm).

Lea este reporte de marketing y después responda lo siguiente:

1. ¿Pueden los datos recopilados aproximarse con una distribución normal?
2. Revise y evalúe las conclusiones hechas por el departamento de marketing de OnCampus! ¿Qué conclusiones son correctas? ¿Cuáles son incorrectas?
3. Si OnCampus! pudiera cambiar su tiempo promedio en cinco minutos, ¿cómo cambiarían las probabilidades?

## REFERENCIAS

1. Gunter, B., “Q-Q Plots”, *Quality Progress* (febrero de 1994), 81-86.
2. Marascuilo, L. A. y M. McSweeney, *Nonparametric and Distribution-Free Methods for the Social Sciences* (Monterey, CA: Brook/Cole, 1977).
3. *Microsoft Excel 2003* (Redmond, WA: Microsoft Corp., 2003).
4. *Minitab for Windows Version 14* (State College, PA: Minitab Inc., 2004).

# Apéndice 6 Uso del software con la distribución normal

## A6.1 EXCEL

### Para probabilidades normales

Abra el archivo **Normal.xls** mostrado en la figura 6.18 de la página 192. Esta hoja de trabajo resuelve los problemas para los ejemplos 6.5 y 6.6. Abra el archivo **Normal Expanded.xls** para solucionar los problemas similares a los ejemplos 6.2 y 6.3, así como los ejemplos 6.5 y 6.6. Ambas hojas de trabajo usan las funciones STANDARDIZE, DISTR.NORM, NORMSINV y NORMINV para calcular probabilidades normales y valores relacionados (vea la sección G.8 para información adicional).

Para adaptar estas hojas de trabajo a otros problemas, cambie los valores de **Media**, **Desviación Estándar**, **Valores X** y **Porcentaje Acumulado** en las celdas sombreadas.

O vea la sección G.8 (Normal) si desea usar el PHStat2 para generar una hoja de trabajo para usted.

### Para una gráfica de probabilidad normal

No hay comandos en Excel que generen directamente una gráfica de probabilidad normal. Si desea que PHSat2 produzca una gráfica de probabilidad normal, consulte la sección G.9.

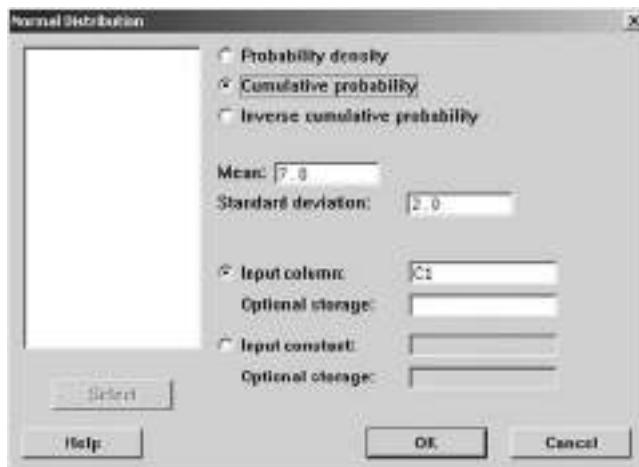
## A6.2 MINITAB

### Uso de Minitab para calcular probabilidades normales

Puede usar Minitab en lugar de la tabla E.2 para calcular probabilidades normales. Para encontrar la probabilidad de que un tiempo de descarga sea menor a 9 segundos con una  $\mu = 7$  y una  $\sigma = 2$ :

1. Ingrese **9** en la primera fila de la columna C1 de una hoja de trabajo en blanco.

2. Seleccione **Calc → Probability Distributions → Normal**.
3. En la ventana de diálogo de la distribución normal (vea la figura A6.1) seleccione el botón de la opción **Cumulative probability**. Ingrese 7 en el cuadro de edición Mean y 2 en el cuadro de edición **Standard deviation**. Seleccione el botón de la opción **Input column** e ingrese **C1** en su cuadro de edición. Dé clic en el botón **OK**. Obtendrá el resultado mostrado en la parte superior de la figura 6.19 de la página 192.



**FIGURA A6.1** Ventana de diálogo de distribución normal de Minitab.

Para encontrar el valor Z correspondiente a un área acumulativa de 0.10,

1. Ingrese **.10** en la fila 1 de la columna C2.
2. Seleccione **Calc → Probability Distributions → Normal**.
3. Seleccione el botón con la opción **Inverse cumulative probability**. Ingrese **7** en el cuadro de edición Mean y **2** en el cuadro de edición **Standard deviation**.
4. Seleccione el botón de la opción **Input column** e ingrese **C2** en el cuadro de edición. Dé clic en el botón **OK**.

Obtendrá el resultado mostrado en la parte inferior de la figura 6.19 de la página 192.

### Uso de Minitab para una gráfica de probabilidad normal

Para construir una gráfica de probabilidad normal en Minitab para los rendimientos en 2003 de los fondos de inversión, abra la hoja de trabajo **MUTUALFUNDS2004.MTV**. Entonces,

1. Seleccione **Graph → Probability Plot**.
2. En el cuadro de diálogo de Probability Plots, seleccione **Single**. Dé clic en el botón **OK**.

3. En el cuadro de diálogo de Probability Plot-Single (vea la figura A6.2), en el cuadro de edición Graph variables, ingrese **C7** o **“Return 2003”**.



**FIGURA A6.2** Ventana de diálogo Probability Plot-Single de Minitab.

4. Dé clic en el botón **Distribution**. En la ventana de diálogo de Probability Plot-Distribution (vea la figura A6.3), seleccione **Normal** en la lista que se despliega hacia abajo de Distribution. Dé clic en el botón **OK** para regresar al cuadro de diálogo de Probability Plot-Single. Dé clic en el botón **OK**.



**FIGURA A6.3** Cuadro de diálogo Probability Plot Distribution de Minitab.



# CAPÍTULO 7

## Distribuciones muestrales

---

**USO DE LA ESTADÍSTICA:** Proceso de empaquetado de cajas de cereal

### 7.1 DISTRIBUCIONES MUESTRALES

- 7.2 DISTRIBUCIÓN MUESTRAL DE LA MEDIA
  - Propiedad de imparcialidad de la media muestral
  - Error estándar de la media
  - Muestreo de poblaciones con distribución normal
  - Muestreo de poblaciones sin distribución normal
    - Teorema del límite central

### 7.3 DISTRIBUCIONES MUESTRALES DE UNA PROPORCIÓN

### 7.4 TIPOS DE MÉTODOS DE MUESTREO PARA ENCUESTAS

- Muestra aleatoria simple
- Muestra sistemática

Muestra estratificada

Muestra de conglomerados

### 7.5 EVALUACIÓN DE LAS VIRTUDES DE UNA ENCUESTA

- Errores de encuesta
- Consideraciones éticas

### A.7 USO DE SOFTWARE PARA LAS DISTRIBUCIONES MUESTRALES

*A7.1 Excel*

*A7.2 Minitab*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- El concepto de distribución muestral
- A calcular probabilidades relacionadas con la media y la proporción muestrales
- La importancia del Teorema del límite central
- A distinguir entre los distintos métodos de muestreo para encuestas

# USO DE LA ESTADÍSTICA



## Proceso de empaquetado de cajas de cereal

En las plantas de Oxford Cereal Company se llenan miles de cajas de cereales en cada turno de ocho horas. Como gerente operativo de una planta, usted se encarga de supervisar la cantidad de cereal que contienen las cajas. Se supone que la media del llenado es de 368 gramos de cereal, como indica la etiqueta del paquete. Por la rapidez del proceso, el peso del cereal varía de una caja a otra, provocando que algunas cajas tengan excedente y otras faltante. Si el proceso no funciona adecuadamente, el peso neto del contenido de las cajas podría diferir mucho de los 368 gramos que indica la etiqueta, hasta resultar inaceptable. Como pesar cada una de las cajas sería un proceso muy lento, costoso e ineficaz, usted debe sacar una muestra de cajas y tomar una decisión con respecto a la probabilidad de que el proceso de llenado de cereales funcione de manera adecuada. Cada vez que selecciona una muestra y pesa las cajas que la componen, calcula la media muestral  $\bar{X}$ . Necesita determinar la probabilidad de que  $\bar{X}$  haya sido extraída de forma aleatoria de una población cuya media poblacional es de 368 gramos. Con base en esta evaluación, tendrá que decidir si conserva, modifica o detiene el proceso.

En el capítulo anterior, usted utilizó la distribución normal para estudiar la distribución de los tiempos de descarga del sitio Web OnCampus! En este capítulo, necesita decidir sobre el proceso de llenado, con base en una *muestra* de cajas de cereal. Aprenderá acerca de las *distribuciones muestrales* y cómo utilizarlas para resolver problemas empresariales. Al igual que en el capítulo anterior, para calcular las probabilidades se utiliza la distribución normal.

## 7.1 DISTRIBUCIONES MUESTRALES

En muchas aplicaciones usted quiere realizar inferencias estadísticas, esto es, utilizar estadísticos calculados a partir de muestras para estimar los valores de los parámetros de la población. En este capítulo aprenderá sobre la media muestral, un estadístico utilizado para estimar la media poblacional (un parámetro). También aprenderá acerca de la proporción muestral, un estadístico que se utiliza para estimar la proporción poblacional (un parámetro). El principal problema al realizar una inferencia estadística radica en obtener conclusiones sobre la población, *no* sobre la muestra. Por ejemplo, una persona que se encarga de realizar encuestas políticas se interesa en los resultados muestrales sólo como mecanismo para estimar la proporción de votos real que recibirá cada uno de los candidatos a partir de la población de votantes. Del mismo modo, como gerente operativo de la Oxford Cereal Company, a usted sólo le interesa utilizar la media muestral calculada a partir de una muestra de cajas de cereal para estimar el peso medio incluido en una población de cajas.

En la práctica, de la población total usted selecciona una muestra aleatoria simple de tamaño predeterminado. Determina qué elementos forman parte de la muestra mediante el uso de un generador de número aleatorio, como una tabla de números aleatorios (vea la sección 7.4 y la tabla E.1), o a través de Excel o Minitab (vea la página 234).

Hipotéticamente, al utilizar un estadístico muestral para estimar un parámetro poblacional, debe examinar *toda* posible muestra que pudiera presentarse. La **distribución muestral** es la distribución de los resultados que se presentan si en realidad se seleccionaron todas las muestras posibles.

## 7.2 DISTRIBUCIÓN MUESTRAL DE LA MEDIA

En el capítulo 3 analizamos varias medidas de tendencia central. Sin duda, la media es la medición de la tendencia central que más se utiliza. Con frecuencia, la media muestral se utiliza para calcular la media poblacional. La **distribución muestral de la media** es la distribución de todas las medias posibles que surgen si en realidad se seleccionaran todas las muestras posibles de cierto tamaño.

### Propiedad de imparcialidad de la media muestral

La media muestral es **imparcial** porque la media de todas las medias muestrales posibles (de una muestra dada con tamaño  $n$ ) es igual a la media poblacional  $\mu$ . Esta propiedad se demuestra por medio de un sencillo ejemplo referente a una población de cuatro asistentes administrativos. Se pide a cada uno de los asistentes que teclee la misma página de un manuscrito. La tabla 7.1 muestra el número de errores.

**TABLA 7.1**

Número de errores cometidos por cada uno de los cuatro asistentes administrativos.

Asistente administrativo	Número de errores
Ana	$X_1 = 3$
Roberto	$X_2 = 2$
Carla	$X_3 = 1$
David	$X_4 = 4$

En la figura 7.1 se muestra esta distribución poblacional.

**FIGURA 7.1**

Número de errores cometidos por una población de cuatro asistentes administrativos.



Cuando se tienen los datos de una población, la media se calcula utilizando la ecuación (7.1):

#### MEDIA POBLACIONAL

La media poblacional es la suma de los valores de la población dividida por el tamaño de la población  $N$ .

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (7.1)$$

Calcule la desviación estándar  $\sigma$  de la población utilizando la ecuación (7.2):

#### DESVIACIÓN ESTÁNDAR POBLACIONAL

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (7.2)$$

Así, para los datos de la tabla 7.1:

$$\mu = \frac{3 + 2 + 1 + 4}{4} = 2.5 \text{ errores}$$

y

$$\sigma = \sqrt{\frac{(3 - 2.5)^2 + (2 - 2.5)^2 + (1 - 2.5)^2 + (4 - 2.5)^2}{4}} = 1.12 \text{ errores}$$

Si a partir de esta población usted selecciona muestras compuestas por dos asistentes administrativos *con reemplazo*, habrá 16 muestras posibles ( $N^n = 4^2 = 16$ ). La tabla 7.2 presenta los resultados de las 16 muestras posibles. Si calcula el promedio de las medias de las 16 muestras, la media de estos valores,  $\mu_{\bar{X}}$ , es igual a 2.5, que es también la media poblacional  $\mu$ .

**TABLA 7.2**

Las 16 muestras de  $n = 2$  asistentes administrativos, de una población de  $N = 4$  asistentes administrativos, al muestrear con reemplazo.

Muestra	Asistentes administrativos	Resultado de la muestra	Media muestral
1	Ana, Ana	3, 3	$\bar{X}_1 = 3$
2	Ana, Roberto	3, 2	$\bar{X}_2 = 2.5$
3	Ana, Carla	3, 1	$\bar{X}_3 = 2$
4	Ana, David	3, 4	$\bar{X}_4 = 3.5$
5	Roberto, Ana	2, 3	$\bar{X}_5 = 2.5$
6	Roberto, Roberto	2, 2	$\bar{X}_6 = 2$
7	Roberto, Carla	2, 1	$\bar{X}_7 = 1.5$
8	Roberto, David	2, 4	$\bar{X}_8 = 3$
9	Carla, Ana	1, 3	$\bar{X}_9 = 2$
10	Carla, Roberto	1, 2	$\bar{X}_{10} = 1.5$
11	Carla, Carla	1, 1	$\bar{X}_{11} = 1$
12	Carla, David	1, 4	$\bar{X}_{12} = 2.5$
13	David, Ana	4, 3	$\bar{X}_{13} = 3.5$
14	David, Roberto	4, 2	$\bar{X}_{14} = 3$
15	David, Carla	4, 1	$\bar{X}_{15} = 2.5$
16	David, David	4, 4	$\bar{X}_{16} = 4$
			$\mu_{\bar{X}} = 2.5$

Puesto que la media de las 16 medias muestrales es igual a la media poblacional, la media muestral es un estimador imparcial de la media poblacional. Por lo tanto, aunque no sepa qué tan cercana está la media muestral de cualquier muestra seleccionada a la media poblacional, al menos estará seguro de que la media de todas las medias muestrales posibles que se pueden seleccionar es igual a la media poblacional.

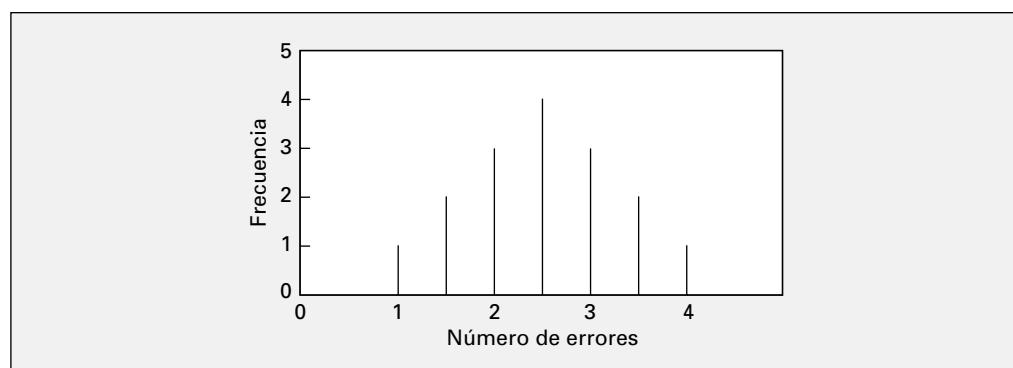
### Error estándar de la media

La figura 7.2 ilustra la variación de la media muestral al seleccionar las 16 muestras posibles. En este pequeño ejemplo, aunque la media muestral varía de una muestra a otra, en función de los asistentes administrativos seleccionados, no lo hace tanto como los valores individuales de la población. El que las medias muestrales varíen menos que los valores individuales es resultado de que cada media muestral promedia todos los valores de la muestra. Una población se compone de resultados individuales que pueden tomar una gran variedad de valores, desde los extremadamente pequeños hasta los muy grandes. No obstante, si una muestra contiene un valor extremo, aunque éste influya sobre la media muestral, el efecto se reduce gracias a que su valor se promedia con todos los demás incluidos en la muestra. Conforme aumenta el tamaño de la muestra, se reduce el efecto de un solo valor extremo, ya que se le promedia con más valores.

**FIGURA 7.2**

Distribución muestral de la media basada en todas las muestras posibles conformadas por dos asistentes administrativos.

Fuente: Datos procedentes de la tabla 7.2.



El valor de la desviación estándar de todas las medias muestrales posibles, llamado **error estándar de la media**, expresa cuánto varía la media muestral entre una muestra y otra. La ecuación (7.3) define al error estándar de la media al hacer muestras *con o sin* reemplazo (vea la página 221) de una población enorme o infinita.

### ERROR ESTÁNDAR DE LA MEDIA

El error estándar de la media  $\sigma_{\bar{X}}$  es igual a la desviación estándar de la población  $\sigma$  dividida por la raíz cuadrada del tamaño de la muestra  $n$ .

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

Por lo tanto, cuando aumenta el tamaño de la muestra, el error estándar de la media se reduce en un factor igual a la raíz cuadrada del tamaño de la muestra.

También es factible utilizar la ecuación (7.3) como aproximación del error estándar de la media cuando la muestra seleccionada sin reemplazo incluye a menos del 5% de la población total. En el ejemplo 7.1 se calcula el error estándar de la media para tal situación.

### EJEMPLO 7.1

### CALCULAR EL ERROR ESTÁNDAR DE LA MEDIA

Volvamos al proceso de llenado de cereal descrito en el escenario “Uso de la estadística” de la página 206. Si de los miles de cajas empacadas durante un turno usted selecciona de manera aleatoria una muestra de 25 cajas sin reemplazo, la muestra contendrá mucho menos del 5% de la población. Si la desviación estándar del proceso de llenado de cereal es de 15 gramos, calcule el error estándar de la media.

**SOLUCIÓN** Al utilizar la ecuación (7.3) mencionada antes, con  $n = 25$  y  $\sigma = 15$ , se determina que el error estándar de la media es

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3$$

La variación de las medias muestrales correspondientes a las muestras de  $n = 25$  es mucho menor que la variación en las cajas de cereal individuales (es decir,  $\sigma_{\bar{X}} = 3$ , mientras que  $\sigma = 15$ ).

### Muestreo de poblaciones con distribución normal

Ahora que ya se explicó el concepto de distribución muestral y que se ha definido el error estándar de la media, ¿qué distribución tendrá la siguiente media muestral  $\bar{X}$ ? Si está muestreando una población que tiene una distribución normal con media  $\mu$  y desviación estándar  $\sigma$ , independientemente del tamaño de la muestra  $n$ , la distribución muestral de la media tendrá una distribución normal con media  $\mu_{\bar{X}} = \mu$  y error estándar de la media  $\sigma_{\bar{X}}$ .

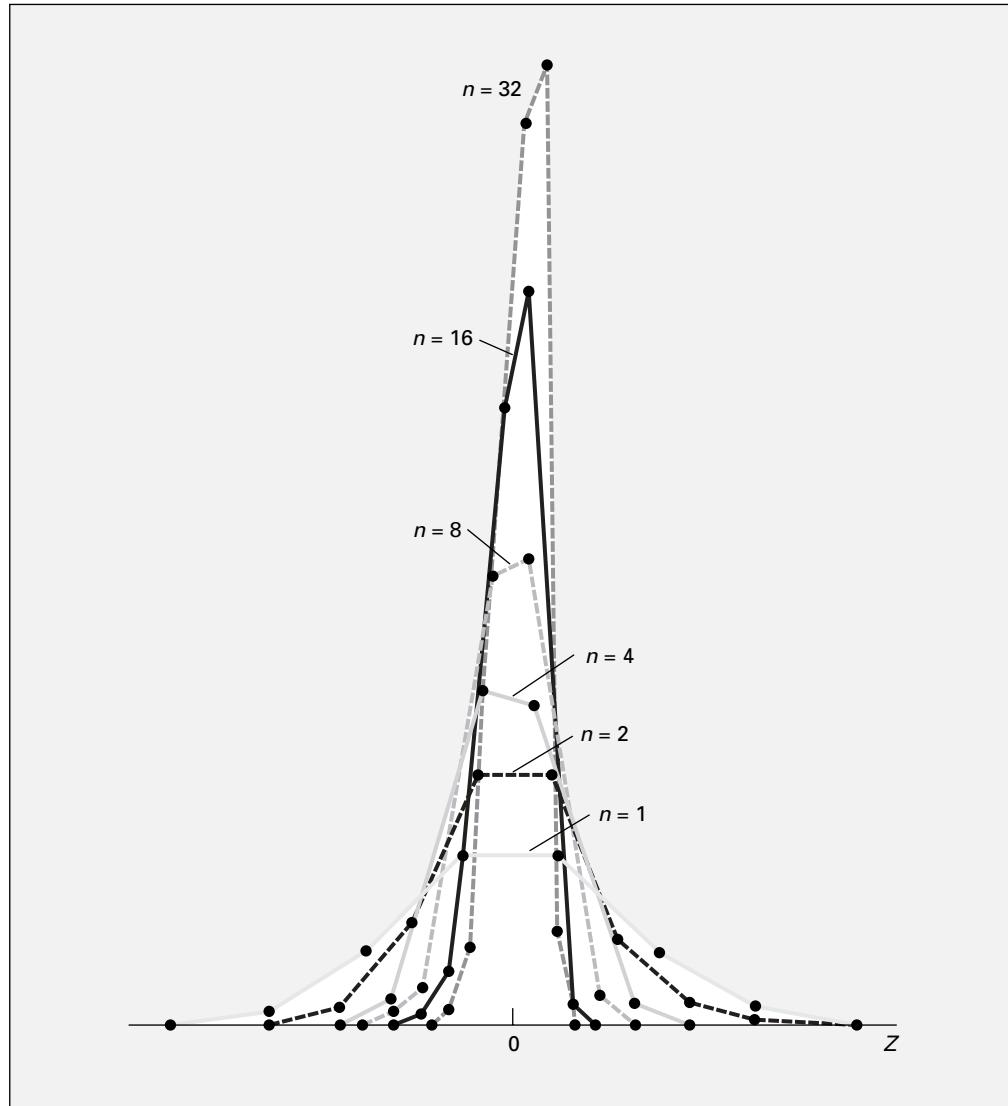
En el caso más simple, si extrae muestras de tamaño  $n = 1$ , cada una de las medias muestrales posibles es un solo valor de la población, puesto que

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_i}{1} = X_i$$

Por lo tanto, si la población tiene una distribución normal con una media  $\mu$  y una desviación estándar  $\sigma$ , entonces la distribución muestral de  $\bar{X}$  para muestras de  $n = 1$  también debe seguir una distribución normal con media  $\mu_{\bar{X}} = \mu$  y error estándar de la media  $\sigma_{\bar{X}} = \sigma/\sqrt{1} = \sigma$ . Además, al aumentar el tamaño de la muestra, la distribución muestral de la media conserva una distribución normal con media  $\mu_{\bar{X}} = \mu$ , pero se reduce el error estándar de la media, de manera que una mayor proporción de las medias muestrales se encuentra más cerca de la media poblacional. La figura 7.3

**FIGURA 7.3**

Distribución muestral de la media de 500 muestras con tamaños de  $n = 1, 2, 4, 8, 16$  y  $32$ , seleccionadas de una población normal.



<sup>1</sup>Recuerde que se seleccionaron "sólo" 500 de un número infinito de muestras, de manera que las distribuciones que se presentan son únicamente aproximaciones de las verdaderas distribuciones.

ilustra esta reducción de la variabilidad en la que se seleccionaron 500 muestras con tamaños de 1, 2, 4, 8, 16 y 32 de manera aleatoria de una población con distribución normal. A partir de los polígonos de la figura 7.3, usted observa que a pesar de que la distribución muestral de la media es aproximadamente<sup>1</sup> normal para cada tamaño de la muestra, las medias muestrales tienen una distribución más estrecha alrededor de la media poblacional conforme aumenta el tamaño de la muestra.

Al examinar el concepto de la distribución muestral de la media, considere de nuevo el escenario "Uso de la estadística" descrito en la página 206. El equipo empaquetador que llena las cajas con 368 gramos de cereal está programado de manera tal que la cantidad de cereal introducido en una caja tenga una distribución normal, con una media de 368 gramos. A partir de la experiencia, se determinó que la desviación estándar poblacional de este proceso de llenado es de 15 gramos.

Si se selecciona de manera aleatoria una muestra 25 cajas de las miles que se llenan durante un día, y se calcula el peso promedio de esta muestra, ¿qué tipo de resultado se espera? Por ejemplo, ¿cree que la media muestral puede ser de 368 gramos? ¿De 200 gramos? ¿De 365 gramos?

La muestra actúa como una representación en miniatura de la población, de tal manera que si los valores de la población tienen una distribución normal, los valores de la muestra deben tener una distribución aproximadamente normal. De esta forma, si la media poblacional es de 368 gramos, la media muestral tiene buenas posibilidades de acercarse a 368 gramos.

¿Cómo se determina la probabilidad de que una muestra de 25 cajas tenga una media inferior a 365 gramos? A partir de la distribución normal (sección 6.2) usted sabe que puede encontrar el área debajo de cualquier valor  $X$  al convertir a unidades estandarizadas  $Z$ .

$$Z = \frac{X - \mu}{\sigma}$$

En los ejemplos de la sección 6.2, se estudió cuánto difiere de la media cualquier valor sencillo  $X$ . Ahora, en el ejemplo de llenado de cajas de cereal, el valor implicado es una media muestral  $\bar{X}$  y usted quiere determinar la probabilidad de que una media muestral sea inferior a 365. De esta manera, al sustituir  $\bar{X}$  por  $X$ ,  $\mu_{\bar{X}}$  por  $\mu$  y  $\sigma_{\bar{X}}$  por  $\sigma$ , se define el valor  $Z$  apropiado en la ecuación (7.4):

### CÁLCULO DE $Z$ PARA LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

El valor  $Z$  es igual a la diferencia que existe entre la media muestral  $\bar{X}$  y la media poblacional  $\mu$ , dividida por el error estándar de la media  $\sigma_{\bar{X}}$ .

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (7.4)$$

Para encontrar el área que se encuentra debajo de 365 gramos, a partir de la ecuación (7.4),

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{365 - 368}{\frac{15}{\sqrt{25}}} = \frac{-3}{3} = -1.00$$

En la tabla E.2 el área correspondiente a  $Z = -1.00$  es 0.1587. Por lo tanto, el 15.87% de todas las muestras posibles con un tamaño de 25 tienen una media muestral inferior a 365 gramos.

Esta afirmación no es lo mismo que decir que cierto porcentaje *individual* de las cajas tendrá menos de 365 gramos de cereal. Calcule ese porcentaje de la siguiente manera:

$$Z = \frac{X - \mu}{\sigma} = \frac{365 - 368}{15} = \frac{-3}{15} = -0.20$$

En la tabla E.2, el área correspondiente a  $Z = -0.20$  es 0.4207. Por lo tanto, cabe esperar que el 42.07% de todas las cajas *individuales* tengan un contenido inferior a 365 gramos. Al comparar esos resultados, observará que se encuentran muchas más *cajas individuales* debajo de los 365 gramos

que *medias muestrales*. Este resultado se explica por el hecho de que cada una de las muestras se compone de 25 valores distintos, algunos pequeños y otros grandes. Al promediar, se diluye la importancia de cualquier valor individual, sobre todo cuando el tamaño de la muestra es grande. De esta forma, la posibilidad de que la media muestral de 25 cajas esté alejada de la media poblacional es menor que la posibilidad de que lo esté una sola caja.

Los ejemplos 7.2 y 7.3 ilustran cómo el uso de distintos tamaños de la muestra influye sobre esos resultados.

### EJEMPLO 7.2

#### EFECTO DEL TAMAÑO DE LA MUESTRA $n$ EN EL CÁLCULO DE $\sigma_{\bar{X}}$

¿Cómo se ve afectado el error estándar de la media al aumentar el tamaño de la muestra de 25 a 100 cajas?

**SOLUCIÓN** Si  $n = 100$  cajas, al utilizar la ecuación (7.3) de la página 209 tenemos:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$$

Al cuadruplicar el tamaño de la muestra de 25 a 100, se reduce a la mitad el error estándar de la media (de 3 a 1.5 gramos). Esto prueba que el uso de una muestra más grande tiene como resultado una menor variabilidad de las medias muestrales entre una muestra y otra.

### EJEMPLO 7.3

#### EFECTO DEL TAMAÑO DE LA MUESTRA $n$ SOBRE EL AGRUPAMIENTO DE LAS MEDIAS EN LA DISTRIBUCIÓN MUESTRAL

En el ejemplo de llenado de cereal, si se selecciona una muestra de 100 cajas, ¿cuál es la probabilidad de que la media muestral sea inferior a 365 gramos?

**SOLUCIÓN** Al utilizar la ecuación (7.4) de la página 211:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{365 - 368}{\frac{15}{\sqrt{100}}} = \frac{-3}{1.5} = -2.00$$

De acuerdo con la tabla E.2, el área menor que  $Z = -2.00$  es 0.0228. Por lo tanto, el 2.28% de las muestras compuestas por 100 cajas tienen una media inferior a 365 gramos, en comparación con el 15.87% de las muestras de 25 cajas.

A veces se requiere encontrar el intervalo que abarca una proporción fija de medias muestrales. Se necesita determinar la distancia arriba y abajo de la media poblacional que abarque un área específica de la curva normal. Al utilizar la ecuación (7.4) de la página 211,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Despejando para  $\bar{X}$  se obtiene la ecuación (7.5).

#### CÁLCULO DE $\bar{X}$ PARA LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

$$\bar{X} = \mu + Z \frac{\sigma}{\sqrt{n}} \quad (7.5)$$

El ejemplo 7.4 ilustra el uso de la ecuación (7.5).

### EJEMPLO 7.4

#### DETERMINAR EL INTERVALO QUE ABARCA UNA PROPORCIÓN FIJA DE LAS MEDIAS MUESTRALES

En el ejemplo de llenado con cereal, encuentre un intervalo alrededor de la media poblacional que abarque el 95% de las medias muestrales a partir de muestras de 25 cajas.

**SOLUCIÓN** Si el 95% de las medias muestrales se encuentra dentro del intervalo, entonces un 5% está fuera de él. Divida el 5% en dos partes iguales de 2.5%. En la tabla E.2, el valor de  $Z$  correspondiente a un área de 0.0250 en la cola inferior de la curva normal es  $-1.96$ , y el valor de  $Z$  correspondiente a un área acumulada de 0.975 (es decir, 0.025 de la cola superior de la curva normal) es  $+1.96$ . El valor menor de  $\bar{X}$  (llamado  $\bar{X}_L$ ) y el valor superior de  $\bar{X}$  (llamado  $\bar{X}_U$ ) se encuentra utilizando la ecuación (7.5):

$$\bar{X}_L = 368 + (-1.96) \frac{15}{\sqrt{25}} = 368 - 5.88 = 362.12$$

$$\bar{X}_U = 368 + (1.96) \frac{15}{\sqrt{25}} = 368 + 5.88 = 373.88$$

Por lo tanto, el 95% de todas las medias muestrales correspondientes a muestras de 25 cajas se encuentra entre 362.12 y 373.88 gramos.

### Muestreo de poblaciones sin distribución normal —Teorema del límite central

Hasta este punto de la sección, ha estudiado la distribución muestral de la media de una población con distribución normal. Sin embargo, se sabe que en muchos casos la población no tiene una distribución normal o no es realista suponer tal distribución. Esta situación se trata mediante un teorema muy importante en estadística, a saber, el teorema del límite central.

#### EL TEOREMA DEL LÍMITE CENTRAL

El **teorema del límite central** dispone que cuando el tamaño de la muestra (es decir, el número de valores en cada muestra) es lo *bastante grande*, la distribución muestral de la media tiene una distribución aproximadamente normal. Esto es válido sin importar la forma de la distribución de los valores individuales en la población.

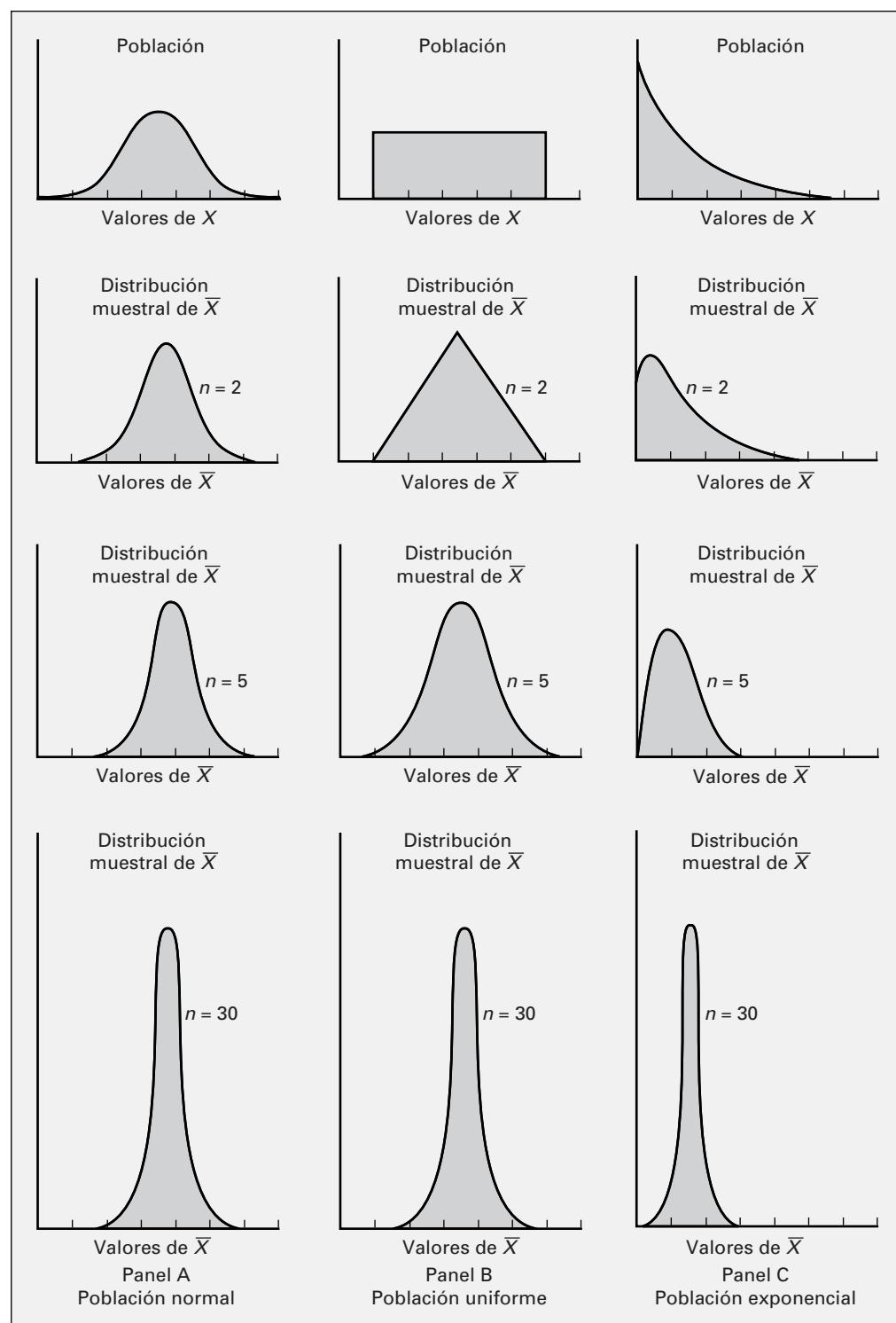
¿Qué tamaño de la muestra es lo bastante grande? A este aspecto se le ha dedicado una gran cantidad de investigación estadística. Por lo general, al utilizar muchas distribuciones de población, los especialistas en estadística han encontrado que cuando el tamaño de la muestra es de por lo menos 30, la distribución muestral de la media es aproximadamente normal. No obstante, si la distribución poblacional tiene una forma aproximada de campana, se aplica el teorema del límite central incluso con tamaños de la muestra menores. Ante el improbable caso de que la distribución sea extremadamente asimétrica o tenga más de una moda, necesitará tamaños de la muestra mayores de 30 para garantizar la normalidad.

La figura 7.4 ilustra la aplicación del teorema del límite central a diferentes poblaciones. Se muestran las distribuciones muestrales de tres distribuciones continuas distintas (normal, uniforme y exponencial) para distintos tamaños de la muestra ( $n = 2, 5, 30$ ).

En el panel A de la figura 7.4 aparece la distribución muestral de la media seleccionada de una población normal. Como ya se mencionó, cuando la población tiene una distribución normal, la distribución muestral de la media está distribuida de manera normal para cualquier tamaño de la muestra. (Usted puede medir la variabilidad utilizando el error estándar de la media, ecuación (7.3) de la página 209.) En virtud de la propiedad de imparcialidad, la media de toda distribución muestral siempre es igual a la media de la población.

**FIGURA 7.4**

Distribución muestral de la media de distintas poblaciones para muestras de  $n = 2, 5$  y  $30$ .



El panel B de la figura 7.4 describe la distribución muestral de una población con distribución uniforme (o rectangular). Cuando se seleccionan muestras con un tamaño  $n = 2$ , hay un efecto de compensación o *restricción central* ya en funcionamiento. Para  $n = 5$ , la distribución muestral tiene forma acampanada y aproximadamente normal. Cuando  $n = 30$ , la distribución muestral se asemeja mucho a una distribución normal. Por lo general, cuanto mayor es el tamaño de la muestra, más se inclinará la distribución muestral a seguir una distribución normal. En todos los casos, la media de cada una de las distribuciones muestrales es igual a la media de la población, y la variabilidad disminuye cuando aumenta el tamaño de la muestra.

El panel C de la figura 7.4 muestra una distribución exponencial. Esta población es extremadamente asimétrica hacia la derecha. Cuando  $n = 2$ , la distribución muestral continúa siendo muy asimétrica hacia la derecha, pero mucho menos que la distribución de la población. Para  $n = 5$ , la distribución muestral es más asimétrica, con sólo una ligera asimetría hacia la derecha. Cuando  $n = 30$ , la distribución muestral tiene una apariencia aproximadamente normal. Una vez más, la media de cada una de las distribuciones muestrales es igual a la media de la población, y la variabilidad disminuye al aumentar el tamaño de la muestra.

Utilizando los resultados surgidos de estas reconocidas distribuciones estadísticas (normal, uniforme y exponencial), se obtienen las siguientes conclusiones con respecto al teorema del límite central.

- Para la mayor parte de las distribuciones poblacionales, sin importar su forma, la distribución muestral de la media tiene una distribución aproximadamente normal cuando se seleccionan muestras de por lo menos 30 elementos.
- Si la distribución poblacional es bastante simétrica, la distribución muestral de la media es aproximadamente normal en muestras tan pequeñas como las de 5 elementos.
- Si la población tiene una distribución normal, la distribución muestral de la media también tiene una distribución normal, independientemente del tamaño de la muestra.

El teorema del límite central es importantísimo al utilizar la inferencia estadística para obtener conclusiones sobre una población. Permite elaborar inferencias sobre la media poblacional sin necesidad de conocer la forma específica de su distribución.

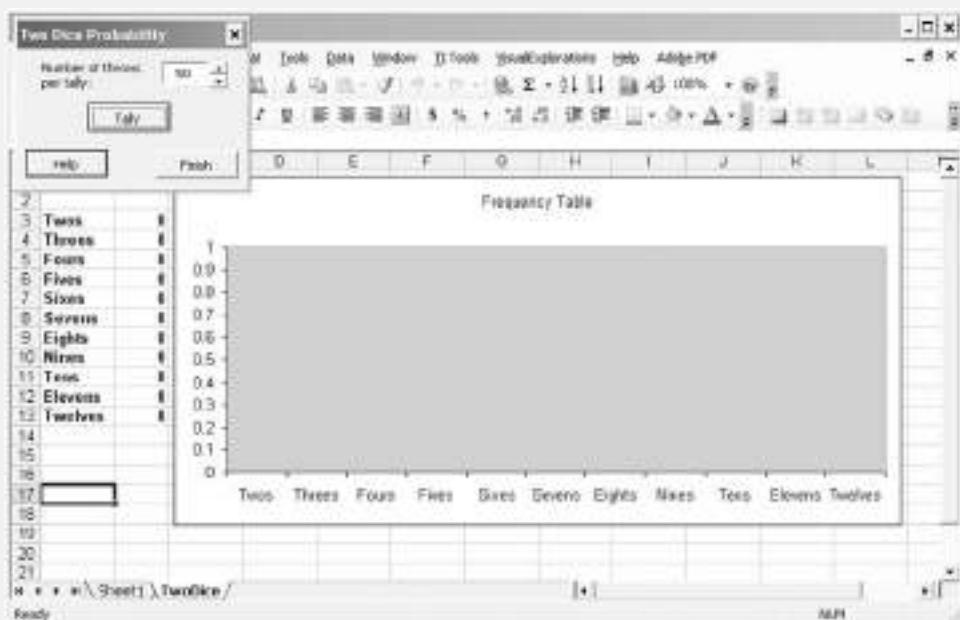
## EXPLORACIONES VISUALES Explorando las distribuciones muestrales

Utilice el procedimiento de Exploraciones Visuales **Probabilidad con dos dados** para observar el efecto que el simulacro de lanzamiento de dados tiene sobre la distribución de frecuencia de la suma de los dos dados. Abra la hoja de trabajo macro Visual Explorations.xls (**Visual Explorations.xls**) y seleccione **VisualExplorations → Two Dice Probability** en la barra de menús de Excel. Este procedimiento genera una hoja de trabajo que contiene una tabla de distribución de frecuencia vacía, un histograma y un

panel de control flotante (vea la ilustración que se presenta abajo).

Dé clic sobre el botón **Tally** para llevar la cuenta de un conjunto de tiros en la tabla y el histograma de la distribución de frecuencia. Como opción, utilice los botones de ajuste para establecer el número de tiros por cuenta (ronda).

Para más información sobre este simulacro, dé clic en el botón **Help**. Cuando termine esta exploración, dé clic en **Finish**.



## PROBLEMAS PARA LA SECCIÓN 7.2

### Aprendizaje básico

**ASISTENCIA de PH Grade** **7.1** Dada una distribución normal con  $\mu = 100$  y  $\sigma = 10$ , si se selecciona una muestra de  $n = 25$ , ¿cuál es la probabilidad de que  $\bar{X}$  sea:

- menor que 95?
- esté entre 95 y 97.5?
- supere 102.2?
- Existe un 65% de posibilidades de que  $\bar{X}$  esté por encima ¿de qué valor?

**ASISTENCIA de PH Grade** **7.2** Dada una distribución normal con  $\mu = 50$  y  $\sigma = 5$ , si se selecciona una muestra de  $n = 100$ , ¿cuál es la probabilidad de que  $\bar{X}$  sea:

- menor que 47?
- esté entre 47 y 49.5?
- supere 51.1?
- Existe un 35% de posibilidades de que  $\bar{X}$  esté por encima ¿de qué valor?

### Aplicación de conceptos

**7.3** Indique cuál es la distribución muestral para muestras de 25 elementos, en cada una de las tres siguientes poblaciones.

- Vales de viáticos para una universidad durante un año académico.
- Registros de ausencia (días de ausencia por año) durante 2004 para los empleados de una gran empresa manufacturera.
- Ventas anuales (en galones) de gasolina sin plomo de las gasolineras ubicadas en una ciudad particular.

**7.4** Los siguientes datos representan el número de días de ausencia al año de una población de seis empleados de una empresa pequeña:

1    3    6    7    9    10

- Suponga que muestrea sin reemplazo, selecciona todas las muestras de  $n = 2$  posibles y construye la distribución muestral de la media. Calcule la media de todas las medias muestrales y la media poblacional. ¿Cómo se denomina a esta propiedad?
- Responda al inciso a) considerando todas las muestras posibles con  $n = 3$ .
- Compare la forma de la distribución muestral de la media de los incisos a) y b). ¿Cuál distribución muestral tiene menor variabilidad? ¿Por qué?
- Suponga ahora que muestrea con reemplazo, responda a los incisos a) a c) y compare los resultados. ¿Cuáles distribuciones muestrales tienen menor variabilidad, las de a) o las de b)? ¿Por qué?

**ASISTENCIA de PH Grade** **7.5** El diámetro de las pelotas de ping-pong fabricadas en una enorme planta tiene una distribución aproximadamente normal, con una media de 1.30 pulgadas y una desviación estándar de 0.04 pulgadas. Si usted selecciona una muestra aleatoria de 16 pelotas de ping-pong.

- ¿Cuál es la distribución muestral de la media?

- ¿Cuál es la probabilidad de que la media muestral sea menor que 1.28 pulgadas?

- ¿Cuál es la probabilidad de que la media muestral se encuentre entre 1.31 y 1.33 pulgadas?

- Existe una probabilidad del 60% de que la media muestral se encuentre entre cuáles dos valores simétricamente distribuidos alrededor de la media poblacional?

**7.6** El Departamento de Comercio de Estados Unidos informó que la mediana del precio de una casa nueva vendida en marzo de 2004 fue de \$201,400, y la media del precio fue de \$260,000 (Michael Schroeder, "New-Home Sales Increase 8.9%, the Biggest Rise in Nine Months", *The Wall Street Journal*, 27 de abril, 2004, A15). Suponga que la desviación estándar de los precios es de \$90,000.

- Si toma muestras de  $n = 2$ , describa la forma de la distribución muestral de  $\bar{X}$ .
- Si toma muestras de  $n = 100$ , describa la forma de la distribución muestral de  $\bar{X}$ .
- Si toma una muestra aleatoria de  $n = 100$ , ¿cuál es la probabilidad de que la media muestral sea menor que \$250,000?

**ASISTENCIA de PH Grade** **7.7** El tiempo dedicado al uso del correo electrónico por sesión tiene una distribución normal, con  $\mu = 8$  minutos y  $\sigma = 2$  minutos. Si se selecciona una muestra aleatoria de 25 sesiones.

- ¿Cuál es la probabilidad de que la media muestral se encuentre entre 7.8 y 8.2 minutos?
- ¿Cuál es la probabilidad de que la media muestral se encuentre entre 7.5 y 8 minutos?
- Si selecciona una muestra aleatoria de 100 sesiones, ¿cuál es la probabilidad de que la media muestral se encuentre entre 7.8 y 8.2 minutos?
- Explique la diferencia en los resultados de los incisos a) y c).

**ASISTENCIA de PH Grade** **7.8** La cantidad de tiempo que un cajero de banco dedica a cada cliente tiene una media poblacional de  $\mu = 3.10$  minutos y una desviación estándar de  $\sigma = 0.40$  minutos. Si se selecciona una muestra aleatoria de 16 clientes.

- ¿Cuál es la probabilidad de que el tiempo medio dedicado a cada cliente sea al menos de 3 minutos?
- Existe un 85% de posibilidades de que la media muestral se encuentre por debajo de cuántos minutos?
- ¿Qué suposición debe hacerse para resolver los incisos a) y b)?
- Si selecciona una muestra aleatoria de 64 clientes, existe un 85% de posibilidades de que la media muestral se encuentre debajo ¿de cuántos minutos?

**7.9** El *New York Times* reportó (Laurie J. Flynn, "Tax Surfing", *The New York Times*, 25 de marzo, 2002, C10) que el tiempo medio necesario para descargar la página inicial del sitio Web del Servicio de Recaudación Interna estadounidense [www.irs.gov](http://www.irs.gov) fue de 0.8 segundos. Suponga que el tiempo de descarga tiene una distribución normal con una desviación estándar de 0.2 segundos. Si se selecciona una muestra aleatoria de 30 tiempos de descarga.

- a. ¿Cuál es la probabilidad de que la media muestral sea menor que 0.75 segundos?
- b. ¿Cuál es la probabilidad de que la media muestral se encuentre entre 0.70 y 0.90 segundos?
- c. La probabilidad de que la media muestral se encuentre entre dos valores simétricamente distribuidos alrededor de la media poblacional es del 80%. ¿Cuáles son esos dos valores?
- d. Existe un 90% de probabilidades de que la media muestral sea menor ¿de qué valor?

**7.10** En el artículo descrito en el problema 7.9, también se informaba que el tiempo medio de descarga del sitio Web de H&R Block [www.hrblock.com](http://www.hrblock.com) fue de 2.5 segundos. Suponga que el

tiempo de descarga de este sitio Web tuvo una distribución normal, con una desviación estándar de 0.5 segundos. Si se selecciona una muestra aleatoria de 30 tiempos de descarga.

- a. ¿Cuál es la probabilidad de que la media muestral sea menor que 2.75 segundos?
- b. ¿Cuál es la probabilidad de que la media muestral se encuentre entre 2.70 y 2.90 segundos?
- c. Hay una probabilidad del 80% de que la media muestral se encuentre entre dos valores simétricamente distribuidos alrededor de la media poblacional. ¿Cuáles son esos valores?
- d. Existe un 90% de probabilidades de que la media muestral sea menor ¿de qué valor?

## 7.3 DISTRIBUCIONES MUESTRALES DE UNA PROPORCIÓN

Considere una variable categórica que cuenta sólo con dos clasificaciones: el cliente prefiere su marca o el cliente prefiere la marca de la competencia. Lo que resulta de interés es la proporción de elementos que forman parte de una de las categorías; por ejemplo, la proporción de clientes que prefieren su marca. La proporción de la población, que se representa por medio de  $\pi$ , es la relación de elementos en toda la población que cuentan con la característica de interés. La proporción muestral, que se representa por medio de  $p$ , es la relación de elementos en la muestra que presentan la característica de interés. La proporción muestral es un estadístico que se utiliza para estimar la proporción poblacional, un parámetro. Para calcular la proporción muestral, se asignan calificaciones de 1 o 0 a los dos resultados posibles, con el fin de representar la presencia o ausencia de la característica. Luego, se suman todas las calificaciones de 1 y 0, y el resultado se divide por el tamaño de la muestra,  $n$ . Por ejemplo, si en una muestra de cinco clientes, tres prefirieron su marca y dos no, hay tres 1 y dos 0. Al sumar los tres 1 y los dos 0, y dividirlos por 5 (el tamaño de la muestra), se obtiene una proporción muestral de 0.60.

### PROPORCIÓN MUESTRAL

$$p = \frac{X}{n} = \frac{\text{número de elementos con la característica de interés}}{\text{tamaño de la muestra}} \quad (7.6)$$

La proporción muestral  $p$  asume valores entre 0 y 1. Si todos los individuos cuentan con la característica, se asigna a cada uno un valor de 1 y  $p$  es igual a 1. Si la mitad de individuos cuentan con la característica, asigne un valor de 1 a una mitad y de 0 a la otra mitad, y  $p$  es igual a 0.5. Si ninguno de los individuos posee la característica, asigne a cada uno una calificación de 0, y  $p$  es igual a 0.

Mientras la media muestral  $\bar{X}$  es un estimador parcial de la media poblacional  $\mu$ , el estadístico  $p$  es un estimador parcial de la proporción poblacional  $\pi$ . Por analogía con la distribución muestral de la media, el **error estándar para la proporción**  $\sigma_p$  se da en la ecuación (7.7).

### ERROR ESTÁNDAR PARA LA PROPORCIÓN

$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}} \quad (7.7)$$

Si usted selecciona todas las muestras posibles de cierto tamaño, la distribución de todas las proporciones muestrales posibles se denomina **distribución muestral de la proporción**. Al realizar muestras con reemplazo de una población finita, la distribución muestral de la proporción sigue la distribución binomial, como se analizó en la sección 5.2. Sin embargo, puede utilizar la distribución

normal para aproximar la distribución binomial cuando  $n\pi$  y  $n(1 - \pi)$  son cada uno de por lo menos 5. En la mayoría de los casos en los que se realizan inferencias con respecto a la proporción, el tamaño de la muestra es lo bastante sustancial como para satisfacer las condiciones necesarias para utilizar la aproximación normal (vea la referencia 1). Por eso, en muchos casos, se utiliza la distribución normal para estimar la distribución muestral de la proporción. Sustituyendo  $p$  por  $\bar{X}$ ,  $\pi$  por  $\mu$  y  $\sqrt{\frac{\pi(1 - \pi)}{n}}$  por  $\frac{\sigma}{\sqrt{n}}$  en la ecuación (7.4) de la página 211, resulta la ecuación (7.8).

### DIFERENCIA ENTRE LA PROPORCIÓN MUESTRAL Y LA PROPORCIÓN POBLACIONAL EN UNIDADES NORMALES ESTANDARIZADAS

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad (7.8)$$

Para ilustrar la distribución muestral de la proporción, suponga que el gerente de la sucursal local de un banco determina que el 40% de todos los clientes tienen varias cuentas en el banco. Si se selecciona una muestra aleatoria de 200 clientes, la probabilidad de que la proporción muestral de clientes con varias cuentas sea menor que 0.30 se calcula como sigue.

Puesto que  $n\pi = 200(0.40) = 80 \geq 5$  y  $n(1 - \pi) = 200(0.60) = 120 \geq 5$ , el tamaño de la muestra tiene la magnitud suficiente como para suponer que la distribución muestral de la proporción tiene una distribución aproximadamente normal. Utilizando la ecuación (7.8),

$$\begin{aligned} Z &= \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \\ &= \frac{0.30 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{200}}} = \frac{-0.10}{\sqrt{\frac{0.24}{200}}} = \frac{-0.10}{0.0346} \\ &= -2.89 \end{aligned}$$

Al emplear la tabla E.2, se observa que el área que queda bajo la curva normal menor que  $Z = -2.89$  es 0.0019. Por lo tanto, la probabilidad de que la proporción muestral sea menor que 0.30 es 0.0019, algo muy poco probable. Esto significa que si la proporción real de éxitos en la población es 0.40, se espera que menos de la quinta parte del 1% de las muestras de  $n = 200$  tenga proporciones muestrales menores que 0.30.

## PROBLEMAS PARA LA SECCIÓN 7.3

### Aprendizaje básico

**ASISTENCIA de PH Grade** **7.11** En una muestra aleatoria de 64 personas, 48 de ellas se clasifican como “éxitosas”. Si la proporción poblacional es 0.70.

- Determine la proporción muestral  $p$  de personas “éxitosas”.
- Determine el error estándar para la proporción.

**ASISTENCIA de PH Grade** **7.12** Se seleccionó una muestra aleatoria de 50 amas de casa para participar en una encuesta telefónica. La pregunta clave que se les planteó fue: “¿Tiene usted o algún miembro de su familia un teléfono celular con mensajes de texto?” De las 50 participantes, 15 respondieron que sí y 35 dijeron que no. Si la proporción de población es de 0.40.

- Encuentre la proporción muestral  $p$  de familiares con teléfonos celulares.
- Encuentre el error estándar para la proporción.

**7.13** Los siguientes datos representan las respuestas (Y para sí y N para no) obtenidas de una muestra de 40 universitarios a la pregunta: “¿Tiene usted actualmente acciones bursátiles de cualquier tipo?”

N N Y N N Y N Y N Y N N Y N Y N N Y  
N Y N N N N Y N N Y Y N N N Y N N Y N N

Si la proporción poblacional es 0.30.

- Encuentre la proporción muestral  $p$  de estudiantes universitarios que poseen acciones bursátiles.
- Encuentre el error estándar para la proporción.

## Aplicación de conceptos

ASISTENCIA  
de PH Grade

AUTO  
Examen

**7.14** Una persona que realiza encuestas políticas efectúa un análisis de resultados muestrales, con objeto de hacer pronósticos sobre la noche de la elección. Suponiendo que se trata de una elección en la que sólo participan dos candidatos, si en la muestra uno de ellos recibe el 55% de los votos, entonces se pronostica que será el ganador de la elección. Si se selecciona una muestra aleatoria de 100 votantes, determine cuál es la probabilidad de que se pronostique un candidato como ganador cuando:

- El porcentaje real de sus votos es del 50.1%.
- El porcentaje real de sus votos es del 60%.
- El porcentaje real de sus votos es del 49% (y en realidad perderá la elección).
- Si se aumenta el tamaño de la muestra a 400, ¿cuáles serían sus respuestas a los incisos a), b) y c)? Analícelo.

ASISTENCIA  
de PH Grade

**7.15** Usted planea realizar un experimento de marketing en el que los estudiantes deben probar una de dos marcas de bebidas gaseosas distintas. Su labor consiste en identificar correctamente cuál es la marca que probaron. Usted selecciona una muestra aleatoria de 200 estudiantes y supone que no cuentan con facultades para distinguir entre ambas marcas. (*Nota:* Si un individuo carece de facultades para distinguir entre los dos bebidas gaseosas, entonces ambas marcas tienen las mismas probabilidades de resultar seleccionadas.)

- ¿Cuál es la probabilidad de que la muestra obtenga entre 50 y 60% de identificaciones correctas?
- Hay una probabilidad del 90% de que el porcentaje muestral se encuentre dentro de ¿cuáles límites simétricos del porcentaje poblacional?
- ¿Cuál es la probabilidad de que el porcentaje muestral de identificaciones correctas sea mayor del 65%?
- ¿Qué es más probable que ocurra: más del 60% de identificaciones correctas en la muestra de 200, o más del 55% de identificaciones correctas en una muestra de 1,000? Explique por qué.

**7.16** Catalyst, una empresa neoyorquina dedicada a la investigación, realizó un estudio sobre las mujeres que ocupan cargos importantes en ambientes corporativos. El estudio concluyó que poco más del 15% de los funcionarios corporativos de las empresas que forman parte de la lista Fortune 500 son mujeres (Carol Hymowitz, "Women Put Noses to the Grindstone, and Miss Opportunities", *The Wall Street Journal*, 3 de febrero, 2004, B1). Suponga que se selecciona una muestra aleatoria de 200 funcionarios corporativos, y que la proporción real de mujeres es de 0.15.

- ¿Cuál es la probabilidad de que en esta muestra menos del 15% de los funcionarios corporativos sean mujeres?
- ¿Cuál es la probabilidad de que en esta muestra entre el 13 y el 17% de los funcionarios corporativos sean mujeres?
- ¿Cuál es la probabilidad de que en esta muestra entre el 10 y el 20% de los funcionarios corporativos sean mujeres?
- Si se hubiese seleccionado una muestra de 100 sujetos, ¿cómo cambiarían sus respuestas de los incisos a) a c)?

**7.17** El programa *Friends* de la NBC fue el espectáculo más popular de TiVo durante la semana del 18 al 24 de abril de 2004.

De acuerdo con el índice Nielsen, el 29.7% de los propietarios de un sistema TiVo en Estados Unidos grabaron o presenciaron en vivo ese programa ("Prime-time Nielsen Ratings", *USA Today*, 28 de abril, 2004, 3D). Suponga que usted selecciona una muestra aleatoria de 50 suscriptores del sistema TiVo.

- ¿Cuál es la probabilidad de que más de la mitad de las personas de la muestra hayan visto o grabado *Friends*?
- ¿Cuál es la probabilidad de que menos del 25% de las personas de la muestra hayan visto o grabado *Friends*?
- Si se hubiera seleccionado una muestra de 500 individuos, ¿cómo cambiarían sus respuestas de los incisos a) y b)?

**7.18** Millones de estadounidenses organizan sus planes de viaje por Internet. De acuerdo con un artículo publicado en *USA Today*, el 77% de los viajeros compran boletos de avión por Internet ("Travelers Head Online", *USA Today Snapshots*, 22 de julio, 2003, A1). Si usted selecciona una muestra aleatoria de 200 viajeros:

- ¿Cuál es la probabilidad de que la muestra contenga entre el 75 y el 80% de viajeros con boleto comprado en Internet?
- Hay una probabilidad del 90% de que el porcentaje muestral se encuentre ¿dentro de cuáles límites simétricos del porcentaje poblacional?
- Hay una probabilidad del 95% de que el porcentaje muestral se encuentre ¿dentro de cuáles límites simétricos del porcentaje poblacional?

**ASISTENCIA  
de PH Grade** **7.19** De acuerdo con la Asociación Nacional de Restaurantes de los Estados Unidos, el 20% de los restaurantes más elegantes han establecido políticas que restringen el uso de teléfonos celulares ("Business Bulletin", *The Wall Street Journal*, 1 de junio, 2000, A1). Si usted selecciona una muestra aleatoria de 100 de los restaurantes más elegantes:

- ¿Cuál es la probabilidad de que la muestra contenga entre el 15 y el 25% de los que han establecido políticas que restringen el uso de teléfonos celulares?
- Hay una probabilidad del 90% de que el porcentaje muestral se encuentre ¿dentro de cuáles límites simétricos del porcentaje poblacional?
- Hay una probabilidad del 95% de que el porcentaje muestral se encuentre ¿dentro de cuáles límites simétricos del porcentaje poblacional?

**7.20** Un artículo (P. Kitchen, "Retirement Plan: To Keep Working", *Newsday*, 24 de septiembre, 2003) analiza los planes de jubilación para los estadounidenses con edades de 50 a 70 años que fueron empleados de tiempo completo o parcial. De los entrevistados, el 29% dijeron que no pensaron trabajar para obtener un salario. Si usted selecciona una muestra aleatoria de 400 estadounidenses con edades de 50 a 70 años que fueron empleados de tiempo completo o parcial:

- ¿Cuál es la probabilidad de que la muestra contenga entre el 25 y el 30% de los que no pensaron en trabajar para obtener un salario?
- Si una muestra actual de 400 estadounidenses con edades de 50 a 70 años que fueron empleados de tiempo completo o parcial contiene un 35% de los que no pensaron en trabajar para obtener un salario, ¿qué se infiere sobre la estimación poblacional del 29%? Explique su respuesta.

- c. Si una muestra actual de 100 estadounidenses con edades de 50 a 70 años que fueron empleados de tiempo completo o parcial contiene un 35% de los que no pensaron en trabajar para obtener un salario, ¿qué se infiere sobre la estimación poblacional del 29%? Explique su respuesta.
- d. Explique las diferencias de los resultados de b) y c).

**7.21** El Servicio de Recaudación Interna de Estados Unidos (IRS por sus siglas en inglés) suspendió las auditorías aleatorias en 1988. En su lugar, instituyó auditorías sobre los rendimientos que su DFS (siglas en inglés para Sistema de Función Discriminante), un complejo sistema computarizado y muy confidencial de análisis, consideró cuestionables. La intención es reducir la proporción de auditorías “sin cambio” (es decir, auditorías que descubren que no se adeudan impuestos adicionales), el IRS sólo audita los rendimientos que el DFS califica como muy cuestionables. La proporción de autorías “sin cambio” ha aumentado con los años, y en la actualidad es de apro-

ximadamente 0.25 (Tom Herman, “Unhappy Returns: IRS Moves to Bring Back Random Audits”, *The Wall Street Journal*, 20 de junio, 2002, A1). Suponga que se selecciona una muestra aleatoria de 100 auditorías. ¿Cuál es la probabilidad de que la muestra tenga

- entre 24 y 26% de auditorías sin cambio?
- entre 20 y 30% de auditorías sin cambio?
- más de 30% de auditorías sin cambio?

**7.22** El IRS anunció que planea reanudar las auditorías totalmente aleatorias el próximo año. Suponga que usted selecciona una muestra aleatoria de 200 auditorías totalmente aleatorias, y que sólo el 10% de todos los rendimientos archivados tienen como resultado auditorías que indican el pago de impuestos adicionales. ¿Cuál es la probabilidad de que la muestra tenga

- entre el 89 y el 91% de auditorías sin cambio?
- entre el 85 y el 95% de auditorías sin cambio?
- más del 95% de auditorías sin cambio?

## 7.4 TIPOS DE MÉTODOS DE MUESTREO PARA ENCUESTAS

En la sección 1.1, se definió la muestra como la parte de la población que se selecciona para realizar el análisis. En lugar de realizar un censo completo de toda la población, los procedimientos de muestreo estadístico se concentran en seleccionar un pequeño grupo representativo de la población. Los resultados de la muestra se utilizan para estimar las características de toda la población. Las tres razones principales para extraer una muestra son:

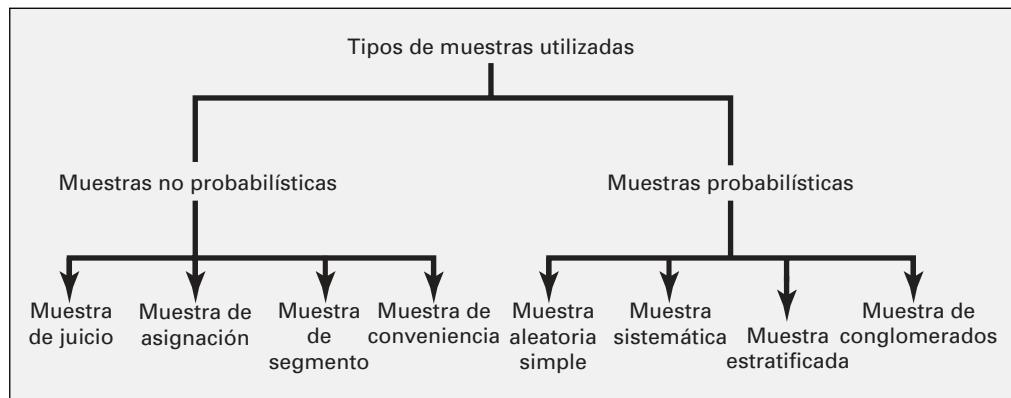
- Una muestra requiere que se le dedique menos tiempo que un censo.
- Es menos costoso administrar una muestra que un censo.
- Una muestra es menos molesta y más práctica de administrar que un censo.

El proceso de muestreo comienza por la definición del marco. El **marco** es una lista de los elementos que constituyen la población. Los marcos son fuentes de datos, como listas, directorios o mapas de la población. Las muestras se extraen de esos marcos. Si los marcos excluyen algunos grupos de la población, los resultados serán inexactos o sesgados. El uso de marcos diferentes para generar los datos podría conducir a conclusiones opuestas.

Una vez seleccionado el marco, se extrae la muestra. Existen dos clases de muestras, que se ilustran en la figura 7.5: las muestras probabilísticas y las no probabilísticas.

**FIGURA 7.5**

Tipos de muestras.



En una **muestra no probabilística**, usted selecciona los elementos o individuos sin conocer sus probabilidades de selección. De tal manera, la teoría desarrollada para el muestreo probabilístico no

se aplica a las muestras no probabilísticas. Un tipo de muestreo no probabilístico es el **muestreo de conveniencia**, en el que los elementos de la muestra se seleccionan sólo con base en el hecho de que son fáciles, económicos o convenientes de muestrear. En algunos casos, los participantes se autoeligen. Por ejemplo, muchas empresas realizan encuestas colocándolas en su página de Internet y brindando la oportunidad a los visitantes de llenar los formatos y enviarlos por la vía electrónica. Las respuestas a estas encuestas ofrecen una gran cantidad de datos rápidamente, pero la muestra se compone de usuarios de Internet autoseleccionados. Para muchos estudios, sólo se dispone de una muestra no probabilística como lo es una muestra de juicio disponible. En una **muestra de juicio** usted recopila las opiniones de expertos en el tema, seleccionados previamente. Otros procedimientos comunes de muestreo no probabilístico son el muestreo de asignación y el de segmento. Estas clases de muestreo se analizan de manera más detallada en libros especializados en métodos de muestreo (vea la referencia 1).

Las muestras no probabilísticas tienen ciertas ventajas como conveniencia, rapidez y menor costo. Sin embargo, su falta de exactitud por el sesgo de la selección y la falta de capacidad de generalización de los resultados opacan estas ventajas. Por esa razón, el uso de los métodos de muestreo no probabilísticos se debe limitar a situaciones en las que se desea obtener aproximaciones flexibles de bajo costo, con el fin de satisfacer la curiosidad sobre algún aspecto en particular, o a estudios de baja escala que preceden a investigaciones más rigurosas.

En una **muestra probabilística**, seleccione los elementos con base en probabilidades conocidas. Siempre que sea posible, debe utilizar métodos de muestreo probabilísticos. Las muestras basadas en estos métodos permiten elaborar inferencias sin sesgo acerca de la población de interés. En la práctica, resulta complicado extraer una muestra probabilística. No obstante, debe tratar de conseguir una muestra probabilística y de reconocer todos los sesgos potenciales que pudieran existir. Los cuatro tipos de muestras probabilísticas de uso más común son: la aleatoria simple, la sistemática, la estratificada y la de conglomerados. Estos métodos de muestreo varían entre sí en su costo, exactitud y complejidad.

## Muestra aleatoria simple

En una **muestra aleatoria simple**, todos los elementos dentro del marco tienen las mismas posibilidades de selección que cualquier otro. Además, cada muestra de un tamaño fijo tiene las mismas posibilidades de selección que cualquier otra muestra del mismo tamaño. El muestreo aleatorio simple es la técnica de muestreo aleatorio más básica y conforma la base de todas las demás técnicas de muestreo aleatorio.

En el muestreo aleatorio simple, la  $n$  se utiliza para representar al tamaño de la muestra y la  $N$  para representar el tamaño del marco. Los elementos del marco se le numeran del 1 a  $N$ . La posibilidad de seleccionar cualquier miembro específico del marco en la primera extracción es  $1/N$ .

Las muestras se seleccionan con reemplazo o sin reemplazo. El **muestreo con reemplazo** implica que, tras seleccionar un elemento, lo devuelve al marco, donde tiene la misma probabilidad de resultar seleccionado de nuevo. Imagine que tiene una urna con  $N$  tarjetas de presentación. En la primera selección, elige la tarjeta de Judith Campos. Usted anota en su registro la información conveniente y devuelve la tarjeta a la urna. Después revuelve las tarjetas y selecciona una segunda tarjeta. En la segunda selección, Judith Campos tiene nuevamente la misma probabilidad de resultar elegida,  $1/N$ . Usted repite el proceso hasta llegar al tamaño de la muestra deseado,  $n$ . Sin embargo, suele ser más conveniente tener una muestra de distintos elementos que permitan la repetición de las medidas del mismo elemento.

**Muestrear sin reemplazo** significa que una vez seleccionado un elemento no se podrá seleccionar de nuevo. La posibilidad de seleccionar cualquier elemento específico del marco en la primera oportunidad, digamos la tarjeta de Judith Campos, es de  $1/N$ . La posibilidad de seleccionar en la segunda tentativa cualquier tarjeta no seleccionada previamente es ahora de 1 dividido por  $N - 1$ . Este proceso continúa hasta completar el tamaño deseado para la muestra.

Independientemente de que realice el muestreo con o sin reemplazo, los métodos “pecera” para seleccionar muestras tienen una gran desventaja: la capacidad de mezclar por completo las tarjetas y sacar la muestra de forma aleatoria. En consecuencia, los métodos pecera no son muy útiles. Necesita utilizar métodos de selección menos engorrosos y más científicos.

Un método consiste en emplear una **tabla de números aleatorios** (vea la tabla E.1) para elegir la muestra. Una tabla de números aleatorios consiste en una serie de dígitos ordenados en una secuencia generada de forma aleatoria (vea la referencia 8). Puesto que el sistema numérico utiliza 10

dígitos (0, 1, 2, ..., 9), la posibilidad que tendrá de generar de forma aleatoria cualquier dígito en particular es **igual** a la probabilidad de generar cualquier otro. Esta probabilidad es de 1 a 10. Por lo tanto, si genera una secuencia de 800 dígitos, podrá esperar que unos 80 tengan el dígito 0, 80 el dígito 1, y así sucesivamente. De hecho, quienes utilizan las tablas de números aleatorios acostumbran probar la aleatoriedad de los dígitos generados antes de utilizarlos. La tabla E.1 satisface todos los criterios de aleatoriedad. Como cada uno de los dígitos o secuencias de dígitos en esta tabla es aleatorio, la tabla se puede leer de forma vertical u horizontal. Los márgenes de la tabla señalan números de líneas y de columnas. Los dígitos se agrupan en secuencias de cinco, con el fin de facilitar la lectura de la tabla.

Al usar una tabla de este tipo en lugar del método de selección de pescera con la finalidad de seleccionar una muestra, es necesario que asigne primero códigos numéricos a los miembros individuales del marco. Así obtiene una muestra aleatoria al leer la tabla de números aleatorios y seleccionar los elementos del marco cuyo número de código concuerda con los dígitos obtenidos de ella. Este proceso de selección de muestras le resultará más comprensible al examinar el ejemplo 7.5.

## EJEMPLO 7.5

### SELECCIÓN DE UNA MUESTRA ALEATORIA UTILIZANDO UNA TABLA DE NÚMEROS ALEATORIOS

Una empresa quiere seleccionar una muestra de 32 trabajadores de tiempo completo, a partir de una población compuesta por 800 empleados de tiempo completo, con el fin de recabar información sobre los gastos referentes al plan de servicios dentales que les brinda. ¿Cómo seleccionaría una muestra aleatoria simple?

**SOLUCIÓN** La empresa supone que no todos responderán la encuesta, por lo que para obtener las 32 respuestas deseadas, es necesario enviar por correo más de 32 formularios. Suponiendo que 8 de cada 10 trabajadores de tiempo completo respondan a la encuesta (es decir, una tasa de respuesta del 80%), decide enviar 40 encuestas.

El marco está formado por una lista con los nombres y números de correo de todos los  $N = 800$  empleados de tiempo completo, tomada de los archivos de personal de la empresa. De esta forma, el marco es un listado completo y exacto de la población. Para seleccionar una muestra aleatoria de 40 empleados de este marco, usted utiliza una tabla de números aleatorios. Puesto que el tamaño de la población (800) es un número de tres dígitos, el código numérico asignado también debe tener tres dígitos, de manera que cada uno de los trabajadores de tiempo completo tenga igual posibilidad de selección. Usted asigna el código 001 al trabajador de tiempo completo que encabeza la lista poblacional, 002 al siguiente empleado de tiempo completo en la lista de empleados, y así sucesivamente, hasta asignar el código 800 al  $N$ -ésimo empleado de la lista. Como  $N = 800$  es el valor de código más grande posible, descarta todos los códigos conformados por las secuencias de tres dígitos mayores que  $N$  (es decir, 801 a 999 y 000).

Para seleccionar la muestra aleatoria simple, usted elige un punto de partida arbitrario en la tabla de números aleatorios. Es válido utilizar el método de cerrar los ojos y señalar con un lápiz algún punto de la tabla de números aleatorios. Suponga que utiliza este procedimiento y selecciona como punto de partida la casilla que ocupa la fila 06 y la columna 05 de la tabla 7.3 (obtenida de la tabla E.1). Aunque puede seguir en cualquier dirección, en este ejemplo lea la tabla de izquierda a derecha, en secuencias de tres dígitos sin saltos.

Al individuo que ocupa el primer sitio en la lista de empleados de tiempo completo le asigna el código 003 en la muestra (fila 06 y columnas 05-07), el segundo individuo tiene el número de código 364 (fila 06 y columnas 08-10), y el tercer individuo tiene el número de código 884. Puesto que el código más elevado para cualquier empleado es 800, descarta este número. Los individuos con números de código 720, 433, 463, 363, 109, 592, 470 y 705 corresponden a los que ocupan los lugares tercero a décimo, respectivamente, en la lista de empleados de tiempo completo.

Usted continúa con el proceso de selección hasta tener la muestra del tamaño necesario, que es de 40 empleados de tiempo completo. Si se trata de un muestreo con reemplazo y durante el proceso de selección se repite cualquier secuencia de código de tres dígitos, incluya de nuevo en la muestra al empleado correspondiente a esa secuencia. Si se trata de un muestreo sin reemplazo, ignore la secuencia código repetida.

**TABLA 7.3**

Uso de una tabla de números aleatorios.

Fila	Columna							
	00000 12345	00001 67890	11111 12345	11112 67890	22222 12345	22223 67890	33333 12345	33334 67890
01	49280	88924	35779	00283	81163	07275	89863	02348
02	61870	41657	07468	08612	98083	97349	20775	45091
03	43898	65923	25078	86129	78496	97653	91550	08078
04	62993	93912	30454	84598	56095	20664	12872	64647
05	33850	58555	51438	85507	71865	79488	76783	31708
<b>Inicio de la selección (fila 06, columna 05)</b>	06	97340	03364	88472	04334	63919	36394	11095
	07	70543	29776	10087	10072	55980	64688	68239
	08	89382	93809	00796	95945	34101	81277	20461
	09	37818	72142	67140	50785	22380	16703	53362
	10	60430	22834	14130	96593	23298	56203	92671
	11	82975	66158	84731	19436	55790	69229	28661
	12	39087	71938	40355	54324	08401	26299	49420
	13	55700	24586	93247	32596	11865	63397	44251
	14	14756	23997	78643	75912	83832	32768	18928
	15	32166	53251	70654	92827	63491	04233	33825
	16	23236	73751	31888	81718	06546	83246	47651
	17	45794	26926	15130	82455	78305	55058	52551
	18	09893	20505	14225	68514	46427	56788	96297
	19	54382	74598	91499	14523	68479	27686	46162
	20	94750	89923	37089	20048	80336	94598	26940
	21	70297	34135	53140	33340	42050	82341	44104
	22	85157	47954	32979	26575	57600	40881	12250
	23	11100	02340	12860	74697	96644	89439	28707
	24	36871	50775	30592	57143	17381	68856	25853
	25	23913	48357	63308	16090	51690	54607	72407
								55538

Fuente: Obtenida de The Rand Corporation. A Million Random Digits with 100,000 Normal Deviates (Glencoe, IL.: The Free Press, 1955) y se incluye completa en la tabla E.1 en el apéndice E de este libro.

## Muestra sistemática

En una **muestra sistemática**, los  $N$  elementos del marco se dividen en  $n$  grupos de  $k$  elementos, donde

$$k = \frac{N}{n}$$

Usted redondea  $k$  al entero más cercano. Para seleccionar una muestra sistemática, elija de forma aleatoria el primer elemento a seleccionar de entre los primeros  $k$  elementos del marco. Luego seleccione de los  $n - 1$  elementos restantes tomando cada  $k$ -ésimo elemento a partir de todo el marco.

Si el marco se compone de una lista de cheques, recibos o facturas prenumerados, es más fácil extraer una muestra sistemática que una muestra aleatoria simple. Una muestra sistemática también es un mecanismo muy conveniente para recabar datos de agendas telefónicas, listas de alumnos y artículos consecutivos que salen de una línea de ensamblaje.

Para extraer una muestra sistemática de  $n = 40$  a partir de la población de  $N = 800$  empleados, divida el marco de 800 en 40 grupos de 20 empleados cada uno. Luego seleccione un número aleatorio de los primeros 20 individuos, e incluya cada vigésimo individuo tras la primera selección de la muestra. Por ejemplo, si el primer número seleccionado es 008, sus selecciones subsiguientes son 028, 048, 068, 088, 108, ..., 768 y 788.

Aunque la muestra aleatoria simple y la muestra sistemática son más sencillas de utilizar, por lo general son menos eficientes que otros métodos más elaborados de muestreo probabilístico. Incluso existen mayores posibilidades de que se presenten sesgos en la selección y falta de representatividad de las características de la población en las muestras sistemáticas que en las muestras aleatorias simples. Si existe algún patrón en el marco, quizás tenga fuertes sesgos de selección. Para superar el problema potencial de una representación de grupos específicos desproporcionada en una muestra, es conveniente utilizar métodos de muestreo estratificados o métodos de muestreo conglomerados.

### Muestra estratificada

En una **muestra estratificada**, primero subdivida  $N$  elementos del marco en subpoblaciones separadas, o **estratos**. Un estrato se define mediante algunas características comunes. Seleccione una muestra aleatoria simple dentro de cada uno de los estratos, y combine los resultados de muestras aleatorias simples distintas. Este método es más eficiente que el muestreo aleatorio simple y que el sistemático, porque garantiza la representación de los elementos a lo largo de toda la población. La homogeneidad de los elementos dentro de cada estrato brinda mayor precisión al estimar los parámetros poblacionales subyacentes.

#### EJEMPLO 7.6

#### SELECCIÓN DE UNA MUESTRA ESTRATIFICADA

Una empresa quiere seleccionar una muestra de 32 empleados de tiempo completo, a partir de una población compuesta de 800 empleados de tiempo completo, con el fin de recabar información sobre los gastos referentes al plan de servicios dentales que les brinda. De todos ellos, el 25% es personal administrativo y el 75% no. ¿Cómo seleccionaría la muestra estratificada, con el fin de que la muestra represente a los trabajadores administrativos en la proporción correcta?

**SOLUCIÓN** Si supone una tasa de respuesta del 80%, necesitará distribuir 40 formularios para obtener las 32 respuestas deseadas. El marco está formado por una lista con los nombres y números de correo de todos los  $N = 800$  empleados de tiempo completo incluidos en los archivos de personal de la empresa. Como el 25% de los empleados de tiempo completo son administrativos, primero divide el marco poblacional en dos estratos: una lista subpoblacional con cada uno de los 200 trabajadores de nivel administrativo y otra lista subpoblacional con cada uno de los 600 trabajadores de tiempo completo no administrativos. Puesto que el primer estrato se compone de una lista con 200 administrativos, le asigna códigos numéricos de tres dígitos del 001 al 200. Como el segundo estrato contiene una lista de 600 trabajadores no administrativos, le asigna códigos numéricos de tres dígitos del 001 al 600.

Para recopilar una muestra estratificada proporcional a las dimensiones de los estratos, se selecciona el 25% de la muestra general de primer estrato y el 75% restante del segundo. Usted toma dos muestras aleatorias simples distintas, cada una basada en un punto de partida aleatorio diferente en la tabla de números aleatorios (tabla E.1). De la primera muestra, selecciona 10 de las personas incluidas en la lista de 200 trabajadores administrativos que componen el primer estrato, y la segunda muestra selecciona 30 de las personas incluidas en la lista de 600 trabajadores no administrativos que componen al segundo estrato. Luego combina los resultados, para reflejar la composición de toda la empresa.

### Muestra de conglomerados

En una **muestra de conglomerados**, divida los  $N$  elementos del marco en varios conglomerados, de tal manera que cada uno sea representativo de toda la población. Luego tome una muestra aleatoria de conglomerados, y estudie todos los elementos de cada conglomerado seleccionado. Los **conglomerados** son designaciones de suceso natural, como países, distritos electorales, cuadras de una ciudad, hogares o territorios de venta.

Con frecuencia, el muestreo de conglomerados tiene una mejor relación costo-eficacia que el muestreo aleatorio simple, sobre todo si la población se distribuye en una región geográfica extensa. Sin embargo, para generar resultados tan precisos como los que se obtienen a partir del muestreo aleatorio simple o del estratificado, a menudo el muestreo de conglomerados necesita una muestra de mayor tamaño. En la referencia 1 encontrará una explicación más detallada de los procesos de muestreo sistemático, estratificado y de conglomerados.

## PROBLEMAS PARA LA SECCIÓN 7.4

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 7.23** Considerando una población que contiene  $N = 902$  individuos, ¿qué número de código asigna usted a
- la primera persona de la lista?
  - la cuadragésima persona de la lista?
  - la última persona de la lista?

- 7.24** Considerando una población de  $N = 902$ , verifique que comenzando en la fila 05 de la tabla de números aleatorios (tabla E.1), sólo necesita seis filas para seleccionar una muestra de  $n = 60$  sin reemplazo.

ASISTENCIA  
de PH Grade

- 7.25** Dada una población de  $N = 93$ , comience en la fila 29 de la tabla de números aleatorios (tabla E.1), y leyendo la fila seleccione una muestra de  $n = 15$ :
- sin reemplazo.
  - con reemplazo.

### Aplicación de conceptos

- 7.26** En un estudio consistente en entrevistas personales a los participantes (en lugar de utilizar encuestas por correo o telefónicas), explique por qué resulta menos práctica una muestra aleatoria simple que cualquier otro de los métodos.

ASISTENCIA  
de PH Grade

- 7.27** Usted quiere seleccionar una muestra aleatoria de  $n = 1$ , a partir de una población de tres elementos (llamados *A*, *B* y *C*). La regla para seleccionar la muestra es: lance una moneda, si cae cara escoja el elemento *A*; si cae cruz, láncela de nuevo; si esta vez cae cara escoja el elemento *B* y si cae cruz, escoja *C*. Explique por qué ésta es una muestra aleatoria pero no una muestra aleatoria simple.

- 7.28** Una población tiene cuatro miembros (llamados *A*, *B*, *C* y *D*). A usted le gustaría extraer una muestra aleatoria de  $n = 2$ , que decide hacer de la siguiente forma: lanza una moneda, si cae cara, la muestra serán los elementos *A* y *B*; si cae cruz, la muestra serán los elementos *C* y *D*. Aunque ésta es una muestra aleatoria, no es una muestra aleatoria simple. Explique por qué. (Si resolvió el problema 7.27, compare el procedimiento ahí descrito con el procedimiento especificado en este problema.)

AUTO  
Examen

- 7.29** El presidente de una universidad con una población de  $N = 4,000$  estudiantes de tiempo completo, solicita al secretario que realice una encuesta para medir la satisfacción de la calidad de vida en el plantel. La siguiente tabla muestra una clasificación por género y grado de los 4,000 estudiantes de tiempo completo inscritos:

Género	Designación de la clase				Total
	1°	2°	3°	4°	
Mujeres	700	520	500	480	2,200
Hombres	560	460	400	380	1,800
Total	1,260	980	900	860	4,000

El secretario pretende obtener una muestra probabilística de  $n = 200$  estudiantes y proyectar los resultados a toda la población de estudiantes de tiempo completo.

- Si el marco disponible en los archivos del secretario es un listado alfabético con los nombres de todos los  $N = 4,000$  estudiantes de tiempo completo inscritos, ¿qué tipo de muestra podrá obtener? Discútalo.
- ¿Cuál es la ventaja de seleccionar una muestra aleatoria simple en el inciso a)?
- ¿Cuál es la ventaja de seleccionar una muestra sistemática en el inciso a)?
- Si el marco disponible a partir de los archivos del secretario es una lista con los nombres de todos los  $N = 4,000$  estudiantes de tiempo completo inscritos, recabada a partir de ocho listas alfabéticas distintas basadas en las divisiones de género y grado que se muestran en la tabla anterior, ¿qué tipo de muestra debe extraer? Discútalo.
- Suponga que cada uno de los  $N = 4,000$  estudiantes de tiempo completo inscritos viven en uno de los 20 dormitorios del campus. Cada dormitorio tiene cuatro pisos, con 50 camas cada uno, por lo que aloja 200 estudiantes. Es política de la institución integrar por completo a sus alumnos por género y grado en cada uno de los pisos de cada dormitorio. Si el secretario recaba un marco por medio de una lista de todos los alumnos que ocupan cada uno de los pisos de cada dormitorio, ¿qué tipo de muestra debe extraer? Discútalo.

- 7.30** En un diario de ventas se conservan facturas prenumbradas. Están foliadas del 0001 al 5,000.

- Comenzando en la fila 16 de la columna 1, y procediendo de forma horizontal en la tabla E.1, seleccione una muestra aleatoria simple de 50 números de factura.
- Seleccione una muestra sistemática de 50 números de factura. Para ello, utilice como punto de partida los números aleatorios de la fila 20 de las columnas 5 a 7.
- ¿Las facturas seleccionadas en el inciso a) son las mismas seleccionadas en el inciso b)? ¿Por qué?

ASISTENCIA  
de PH Grade

- 7.31** Suponga que 5,000 facturas se separan en cuatro estratos. El estrato uno contiene 50 facturas, el estrato 2 tiene 500, el estrato 3 contiene 1,000 y el estrato cuatro tiene 3,450 facturas. Se requiere una muestra de 500 facturas.

- ¿Qué tipo de muestreo debe realizar? ¿Por qué?
- Explique cómo realizaría el muestreo, de acuerdo con el método establecido en el inciso a)?
- ¿Por qué el muestreo establecido en el inciso a) no es el muestreo simple?

## 7.5 EVALUACIÓN DE LAS VIRTUDES DE UNA ENCUESTA

Temprano todos los días usted lee o escucha sobre los resultados de encuestas de opinión o investigaciones en periódicos, Internet, radio o televisión. Para identificar las encuestas que carecen de objetividad o credibilidad, debe evaluar de manera crítica lo que lee o escucha, examinando las virtudes de la encuesta. Primero, habrá que evaluar los propósitos de la encuesta, por qué y quién la realizó. Una encuesta o investigación de opinión realizada para satisfacer una curiosidad es esencialmente un entretenimiento. Más que un medio para llegar a un fin, sus resultados son un fin en sí mismos. Usted se debe mostrar escéptico ante este tipo de encuestas, porque el resultado no tendrá utilidad posterior.

El segundo paso al evaluar las virtudes de una encuesta consiste en determinar si se basó en una muestra probabilística o no probabilística (como se analiza en la sección 7.4). Recuerde que la única forma de hacer inferencias estadísticas correctas a partir de una muestra de la población, es mediante el uso de una muestra probabilística. Las encuestas que utilizan métodos de muestreo no probabilísticos están sujetas a sesgos severos, quizás involuntarios, que generaran resultados irrelevantes, como se ilustra en el ejemplo siguiente, tomado de las elecciones presidenciales estadounidenses celebradas en 1948.

En 1948, las principales organizaciones de encuestas pronosticaron que el resultado de las elecciones presidenciales estadounidenses, en las que competían Harry S. Truman, presidente en funciones que buscaba la reelección, y Thomas E. Dewey, entonces gobernador de Nueva York, favorecerían a Dewey. El *Chicago Tribune* confió tanto en los pronósticos de las encuestas que imprimió su primera edición con base en ellos en lugar de esperar al conteo de los sufragios.

El desconcertado periódico y los encuestadores en los que había confiado tuvieron mucho que explicar. ¿Cómo era posible que los encuestadores estuvieran tan equivocados? Tratando de encontrar el origen del error, las organizaciones encuestadoras descubrieron que todo se debía al uso de un método de muestreo no probabilístico (vea la referencia 7). En consecuencia, esas organizaciones adoptaron los métodos de muestreo probabilísticos para las elecciones subsiguientes.

### Errores de encuesta

Aun cuando las encuestas utilizan métodos de muestreo probabilísticos, están sujetas a errores potenciales. Cuatro tipos de errores de encuesta son:

- Error de cobertura.
- Error de no respuesta.
- Error de muestreo.
- Error de medición.

Un buen diseño de encuesta de investigación pretende reducir al mínimo estos distintos errores, con frecuencia a un costo considerable.

**Error de cobertura** La clave para hacer una selección apropiada de la muestra es partir de un marco adecuado. Recuerde, el marco es una lista actualizada de todos los elementos, a partir de la cual podrá seleccionar la muestra. El **error de cobertura** se presenta cuando se excluyen del marco ciertos grupos de elementos, de tal manera que no tienen posibilidad de resultar seleccionados como parte de la muestra. El error de cobertura provoca un **sesgo de selección**. Si el marco es inadecuado porque no se incluyeron de manera apropiada ciertos grupos de elementos, toda muestra probabilística aleatoria seleccionada dará una estimación de las características del marco, no de la población *real*.

**Error de no respuesta** No todas las personas están dispuestas a responder una encuesta. De hecho, las investigaciones han descubierto que los individuos de las clases económicas más altas y más bajas suelen responder a las encuestas con menos frecuencia que las personas de clase media. El **error de no respuesta** surge a partir de la omisión al recabar datos de todos los elementos de la muestra, y tiene como resultado un **sesgo de no respuesta**. Puesto que por regla general no es posible suponer que las personas que declinan responder a la encuesta son semejantes a las que sí lo hacen, es necesario hacer un seguimiento de la falta de respuestas tras un periodo de tiempo espe-

cificado. Usted debe realizar varios intentos para convencer a tales individuos de que respondan la encuesta. Luego, las respuestas de seguimiento se comparan con las respuestas iniciales, para efectuar inferencias válidas a partir de la encuesta (referencia 1).

El modo de respuesta utilizado influye en la tasa de respuesta. Las entrevistas personales y telefónicas suelen producir una tasa de respuesta mayor que las encuestas por correo, pero a mayor costo. A continuación se muestra un famoso ejemplo de error de cobertura y sesgo de no respuesta.

En 1936, la revista *Literary Digest* pronosticó que el gobernador de Kansas, Alf Landon, recibiría el 57% de los votos en las elecciones presidenciales estadounidenses, y derrotaría de forma abrumadora al presidente Franklin D. Roosevelt, que se postulaba para un segundo periodo. Sin embargo, Landon fue fácilmente derrotado al recibir sólo el 38% de los votos. Nunca antes había ocurrido que una revista cometiera un error semejante con respecto a una encuesta tan importante. En consecuencia, tal pronóstico arruinó la credibilidad del público en la revista, lo que, con el tiempo, derivó en bancarrota. En *Literary Digest* creyeron haber hecho todo de manera correcta. Basaron su pronóstico en una muestra de gran tamaño, 2.4 millones de participantes en una encuesta enviada a 10 millones de votantes registrados. ¿Cuál fue el error? Existen dos respuestas: sesgo de selección y sesgo de no respuesta.

Para comprender el papel que desempeña el sesgo de selección, es necesario conocer ciertos antecedentes históricos. En 1936, Estados Unidos todavía sufrió los efectos de la Gran Depresión. Sin tomar esto en cuenta, el *Literary Digest* recopiló su marco de fuentes tales como directorios telefónicos, listas de miembros de clubes, suscriptores y registros de automóvil (referencia 7). Inadvertidamente, seleccionaron un marco compuesto primordialmente por gente acomodada que excluyó a la mayor parte de la población votante, quienes durante la Gran Depresión no podían costear teléfonos, afiliación a clubes, suscripciones ni automóviles. De esta manera, el 57% de votos que se estimó para Landon pudo haber sido muy cercano al marco, pero en realidad no lo fue para la población estadounidense total.

El error de no respuesta generó un posible sesgo, ya que la gran muestra de 10 millones de votantes registrados obtuvo sólo 2.4 millones de respuestas. Una tasa de respuesta de sólo un 24% está muy lejos de producir estimaciones exactas de los parámetros poblacionales, y no cuenta con mecanismos para garantizar que los 7.6 millones de individuos que no respondieron tengan una opinión similar. Sin embargo, el problema causado por el sesgo de no respuesta resulta secundario en comparación con el sesgo de selección. Incluso si hubieran respondido los 10 millones de votantes registrados en la muestra, esto no hubiera compensado el hecho de que la composición del marco difería de manera sustancial de la población votante real.

**Error de muestreo** Existen tres razones primordiales por las que usted selecciona una muestra en vez de realizar un censo completo: es más conveniente, menos costoso y más eficiente. Sin embargo, la casualidad dicta cuáles individuos o elementos serán o no incluidos en la muestra. El **error de muestreo** refleja la heterogeneidad o “diferencia de posibilidad” entre una muestra y otra, con base en la probabilidad de que determinados individuos o elementos sean seleccionados en muestras particulares.

Cuando usted lee los resultados de encuestas en periódicos o revistas, con frecuencia encontrará un enunciado relacionado con el margen de error o la precisión. Por ejemplo, “se considera que los resultados de esta encuesta están a  $\pm 4$  puntos porcentuales del valor real”. Este margen de error representa al error de muestreo. El error de muestreo se reduce extrayendo muestras de mayor tamaño, pero esto también aumenta el costo de la encuesta.

**Error de medición** En la práctica de una buena investigación, usted diseña un cuestionario con intención de recabar información significativa. Pero aquí se enfrenta a un dilema: es más fácil hablar de medidas significativas que obtenerlas. Considere el siguiente proverbio:

*Un hombre con un reloj siempre sabe qué hora es;  
Un hombre con dos relojes siempre indaga, para identificar la hora correcta;  
Un hombre con diez relojes siempre recuerda lo difícil que es medir el tiempo.*

Por desgracia, el proceso de obtener una medición con frecuencia está regido por lo que resulta conveniente y no por lo que es necesario. A menudo las mediciones son en realidad sólo una representación de las que se desean. Se ha observado mucho al error de medición que surge como resul-

tado de la fragilidad de la redacción (referencia 3). Una pregunta debe ser clara y precisa. Además, para evitar las *preguntas tendenciosas*, es necesario plantearlas de manera neutral.

Existen tres causas de **errores de medición**: la redacción ambigua de las preguntas, el efecto halo y el error de la encuesta. Como un ejemplo de redacción ambigua, en noviembre de 1993 el Departamento del Trabajo de Estados Unidos informó que la tasa de desempleo en ese país se había subestimado por más de una década, como resultado de la mala redacción del cuestionario en la Encuesta Actual de Población. De manera específica, la redacción provocó un conteo significativamente menor de las mujeres en la fuerza de trabajo. Puesto que las tasas de desempleo se encuentran ligadas a programas de asistencia social como los seguros de desempleo estatales, era imperativo que los investigadores gubernamentales rectificaran la situación corrigiendo la redacción del cuestionario.

El “efecto halo” se presenta cuando el encuestado se siente obligado a complacer al entrevistador. La adecuada capacitación del entrevistador logra reducir al mínimo el efecto halo.

El error del encuestado se presenta como consecuencia del exceso o de la falta de interés del encuestado. Hay dos formas de reducir al mínimo este error: **1.** analizando cuidadosamente los datos e interrogando de nuevo a los individuos cuya respuestas parecen inusuales, y **2.** estableciendo un programa de reentrevistas aleatorias, con la finalidad de determinar la confiabilidad de las respuestas.



## Consideraciones éticas

Las consideraciones éticas surgen en relación con los cuatro tipos de errores potenciales que pueden surgir al diseñar encuestas que utilizan muestras probabilísticas: error de cobertura, error de no respuesta, error de muestreo y error de medición. El error de cobertura da como resultado un sesgo en la selección, y se convierte en un problema ético si se excluye del marco a individuos o grupos específicos, en forma deliberada, de manera que los resultados de la encuesta se inclinen hacia una posición más favorable para el interesado. El error de no respuesta puede conducir al sesgo de no respuesta, y se convierte en un problema ético si el interesado diseña deliberadamente la encuesta de tal manera que se reduzca a la posibilidad de que la respondan grupos o individuos específicos. El error de muestreo se convierte en un problema ético si al mostrar los resultados se omite deliberadamente hacer referencia al tamaño de la muestra y al margen de error, de manera que el interesado promueva un punto de vista que, de otra forma, no sería verdaderamente significativo. El error de medición se convierte en un problema ético en alguna de las siguientes tres situaciones: **1.** cuando el interesado en la encuesta elige preguntas tendenciosas que dirigen las respuestas en una dirección específica; **2.** cuando el entrevistador, mediante ademanes y tonos, provoca deliberadamente un efecto halo o, de cualquier otra forma, dirige las respuestas en una dirección específica; **3.** cuando el encuestado, desdeñando el proceso de encuestas, proporciona información falsa de manera intencional.

También surgen algunos problemas éticos cuando se utilizan los resultados de las muestras no probabilísticas para sacar conclusiones acerca de toda la población. Cuando se utiliza un método de muestreo no probabilístico, es necesario explicar los procedimientos de muestreo y asentar que los resultados no deben generalizarse más allá de la muestra.

## PROBLEMAS PARA LA SECCIÓN 7.5

### Aplicación de conceptos

**7.32** “Una encuesta señala que la gran mayoría de los estudiantes universitarios son dueños de su propia computadora personal.” ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

**7.33** Se selecciona una muestra aleatoria simple de  $n = 300$  empleados de tiempo completo a partir de una lista de la empresa que contiene los nombres de todos los  $N = 5,000$  empleados de tiempo completo, con el fin de evaluar la satisfacción con el trabajo.

- a. Encuentre un ejemplo de posible error de cobertura.
- b. Encuentre un ejemplo de posible error de no respuesta.

- c. Encuentre un ejemplo de posible error de muestreo.
- d. Encuentre un ejemplo de posible error de medición.

**7.34** De acuerdo con una encuesta realizada a 1,000 suscriptores de AOL (Harry Berkowitz, “Screen Name Loyalty”, *Newsday*, 1 de diciembre, 2002, A42), el 92% de los suscriptores de AOL respondieron: “No quiero cambiar mi dirección de correo electrónico”, como razón para continuar con el servicio. ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

**7.35** Forrester Research Inc. realizó una encuesta entre compradores en línea (Michael Totty, “The Masses Have Arrived”, *The Wall Street Journal*, 27 de enero, 2003, R8).

De quienes han realizado compras en línea por menos de un año, el 39% tiene licenciatura, el 57% está conformado por mujeres, y su ingreso anual medio es de \$52,300. ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

**7.36** De acuerdo con un sondeo realizado por Maritz entre 1,004 conductores adultos (“Snapshots”, *USA Today*, 23 de octubre, 2002), el 45% admitió ingerir alimentos o bebidas de manera frecuente u ocasional, y el 36% admitió hablar por teléfono celular. ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

**7.37** De acuerdo con una encuesta realizada por MessageOne entre personas que tienen correo electrónico en su trabajo, casi 4 de cada 10 indicaron que no pueden vivir sin él (Anne R. Ca-

rey y Chad Palmer, “Snapshots”, *USA Today*, 14 de enero, 2004, A1). De manera más específica, el 37% dijo “no poder vivir sin él”, el 26% dijo que el correo electrónico era “importante”, el 19% considera que “no es esencial”, el 13% contestó “no lo utilizo” y el 5% afirmó que “no es importante”. ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

**7.38** ¿Qué desean quienes cenan en un restaurante? En una encuesta realizada por Caravan para la IHOP, el 56% respondió que quería “alimentos estupendos”. Otras respuestas fueron: “precios razonables”, el 22%; “ambiente” el 11%; “servicio rápido”, el 8%; y “no sé”, el 3% (Darryl Haralson y Jeff Dionise, “Ssnapshots”, *USA Today*, 16 de enero, 2004, A1). ¿Qué información le gustaría conocer antes de aceptar los resultados de la encuesta?

## RESUMEN

En este capítulo se estudió la distribución muestral de la media muestral, el teorema del límite central y la distribución muestral de la proporción de la muestra. Usted aprendió que la media muestral es un estimador sin sesgo de la media poblacional y que la proporción muestral es un estimador sin sesgo de la proporción poblacional. Al observar el peso medio de una muestra de cajas de cereal llenadas por la Oxford Ce-

real Company, usted pudo obtener conclusiones relativas al peso medio de la población de cajas de cereal. También estudió cuatro métodos comunes de muestreo para encuesta: aleatorio simple, sistemático, estratificado y de conglomerados. En los siguientes cuatro capítulos se analizan las técnicas más utilizadas por la inferencia estadística, los intervalos de confianza y las pruebas de hipótesis.

## FÓRMULAS IMPORTANTES

### Media poblacional

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad (7.1)$$

### Desviación estándar poblacional

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} \quad (7.2)$$

### Error estándar de la media

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

### Cálculo de Z para la distribución muestral de la media

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (7.4)$$

### Cálculo de $\bar{X}$ para la distribución muestral de la media

$$\bar{X} = \mu + Z \frac{\sigma}{\sqrt{n}} \quad (7.5)$$

### Proporción muestral

$$p = \frac{X}{n} \quad (7.6)$$

### Error estándar para la proporción

$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}} \quad (7.7)$$

### Cálculo de Z para la distribución muestral de la proporción

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad (7.8)$$

## CONCEPTOS CLAVE

conglomerados 224	error estándar para la proporción 217	muestra probabilística 221
distribución muestral 206	errores de medición 228	muestra sistemática 223
distribución muestral de la media 207	estratos 224	muestrear sin reemplazo 221
distribución muestral de la proporción 217	imparcial 207	muestreo con reemplazo 221
error de cobertura 226	marco 220	muestreo de conveniencia 221
error de muestreo 227	muestra aleatoria simple 221	sesgo de no respuesta 226
error de no respuesta 226	muestra de conglomerados 224	sesgo de selección 226
error estándar de la media 209	muestra de juicio 221	sin sesgo 207
	muestra estratificada 224	tabla de números aleatorios 221
	muestra no probabilística 220	teorema del límite central 213

## PROBLEMAS DE REPASO

### Revisión de su aprendizaje

**7.39** ¿Por qué la media muestral es un estimador sin sesgo de la media poblacional?

**7.40** ¿Por qué se reduce el error estándar de la media conforme aumenta el tamaño de la muestra  $n$ ?

**7.41** ¿Por qué para una muestra con el tamaño suficientemente grande, la distribución muestral de la media sigue una distribución normal a pesar de que la población quizás no tenga una distribución normal?

**7.42** ¿Cuál es la diferencia entre una distribución de probabilidad y una distribución muestral?

**7.43** ¿En qué circunstancias la distribución muestral de la proporción sigue una distribución aproximadamente normal?

**7.44** ¿Cuál es la diferencia entre el muestreo probabilístico y el no probabilístico?

**7.45** ¿Cuáles son algunos de los problemas que se presentan al utilizar métodos de “pecera” al seleccionar una muestra aleatoria simple?

**7.46** ¿Cuál es la diferencia entre muestreo *con* reemplazo y *sin* reemplazo?

**7.47** ¿Cuál es la diferencia que existe entre una muestra aleatoria simple y una muestra sistemática?

**7.48** ¿Cuál es la diferencia que existe entre una muestra aleatoria simple y una muestra estratificada?

**7.49** ¿Cuál es la diferencia que existe entre una muestra estratificada y una muestra de conglomerados?

### Aplicación de conceptos

**7.50** Una máquina de coser industrial utiliza cojinetes que deben tener un diámetro de 0.75 pulgadas. Los límites inferior y superior especificados dentro de los que los cojinetes pueden operar son 0.74 y 0.76 pulgadas, respectivamente. La experien-

cia previa señala que el diámetro real de los cojinetes tiene una distribución aproximadamente normal, con una media de 0.753 pulgadas y una desviación estándar de 0.004 pulgadas. Si se selecciona una muestra aleatoria de 25 cojinetes, ¿cuál es la probabilidad de que la media muestral se encuentre

- entre el objetivo y la media poblacional de 0.753?
- entre el límite inferior de las especificaciones y el objetivo?
- sobre el límite superior a las especificaciones?
- bajo el límite inferior de las especificaciones?
- Existe un 93% de probabilidad de que el diámetro medio muestral sea mayor ¿de qué valor?

**7.51** La cantidad de llenado de las botellas de una bebida gaseosa tiene una distribución normal, con una media de 2.0 litros y una desviación estándar de 0.05 litros. Si se selecciona una muestra aleatoria de 25 botellas, ¿cuál es la probabilidad de que la media muestral esté

- entre 1.99 y 2.0 litros?
- bajo 1.98 litros?
- sobre 2.01 litros?
- Existe una probabilidad del 99% de que la media muestral contenga por lo menos ¿cuánta bebida gaseosa?
- Hay una probabilidad del 99% de que la media muestral contenga valor que se encuentra entre ¿cuáles dos valores (distribuidos simétricamente alrededor de la media)?

**7.52** Un productor de jugo adquiere todas sus naranjas de una enorme plantación que tiene una variedad de naranja. La cantidad de jugo extraído de cada una de estas naranjas tiene una distribución aproximadamente normal, con una media de 4.70 onzas y una desviación estándar de 0.40 onzas. Suponga que selecciona una muestra aleatoria de 25 naranjas:

- ¿Cuál es la probabilidad de que la media muestral sea de por lo menos 4.60 onzas?
- Hay una probabilidad del 70% de que la media muestral se encuentre entre dos valores simétricamente distribuidos alrededor de la media poblacional. ¿Cuáles son esos valores?
- Existe un 77% de probabilidad de que la media muestral esté por encima de ¿cuál valor?

**7.53** DiGiorno's tiene en la televisión algunos de los comerciales más creativos y agradables de pizzas congeladas. La sección revisora de publicidad del *USA Today* asegura que al 20% de los espectadores les gustan "mucho" los anuncios (Theresa Howard, "DiGiorno Campaign Delivers Major Sales", [www.usatoday.com](http://www.usatoday.com), 1 de abril, 2002). Suponga que los anuncios se exponen a una muestra de 400 espectadores de televisión. ¿Cuál es la probabilidad de que la muestra tenga entre

- 18 y 22% de personas a quienes los anuncios les gustan "mucho"?
- 16 y 24% de personas a quienes los anuncios les gustan "mucho"?
- 14 y 26% de personas a quienes los anuncios les gustan "mucho"?
- 12 y 28% de personas a quienes los anuncios les gustan "mucho"?

**7.54** Los fondos de inversión reportaron ganancias moderadas durante el primer trimestre de 2004. Los fondos de participación diversificada estadounidenses, una gran bolsa de acciones compuesta por una amplia variedad de empresas, obtuvo un rendimiento del 2.98% (Michael J. Martines, "Mutual-fund Returns Minimal in First Quarter", [Cincinnati.com](http://www.cincinnati.com), 3 de abril, 2004). Suponga que los rendimientos de los fondos estadounidenses se distribuyeran como una variable aleatoria normal, con una media de 2.98 y una desviación estándar de 4. Si usted seleccionara una muestra aleatoria de 10 fondos de esta población, ¿cuál es la probabilidad de que la muestra tenga un rendimiento medio

- menor que 0; es decir, pérdidas?
- entre 0 y 6?
- mayor que 10?

**7.55** Los fondos de inversión reportaron ganancias moderadas durante el primer trimestre de 2004. Los fondos internacionales, que históricamente han sido más volátiles que los fondos estadounidenses, obtuvieron un rendimiento del 5.13% (Michael J. Martines, "Mutual-fund Returns Minimal in First Quarter", [Cincinnati.com](http://www.cincinnati.com), 3 de abril, 2004). Suponga que los rendimientos de los fondos internacionales se distribuyeron como una variable aleatoria normal, con una media de 5.13 y una desviación estándar de 6. Si usted seleccionara un fondo individual de esta población, ¿cuál es la probabilidad de que tenga un rendimiento

- menor que 0; es decir, pérdidas?
- entre 0 y 6?
- mayor que 10?

Si usted seleccionara una muestra aleatoria de 10 fondos de esta población, ¿cuál es la probabilidad de que la muestra tenga un rendimiento medio

- menor que 0; es decir, pérdidas?
- entre 0 y 6?
- mayor que 10?
- Compare sus respuestas en los incisos d) a f) con las de los incisos a) a c).
- Compare sus respuestas en los incisos d) a f) con las de los incisos a) a c) del problema 7.54.

**7.56** La realización de encuestas políticas se ha valido tradicionalmente de entrevistas telefónicas. Los investigadores de Harris Black International Ltd., argumentan que las encuestas

por Internet son más económicas, más rápidas, y que obtienen mayores tasas de respuesta que las telefónicas. Los críticos están preocupados por la confiabilidad de este método (*The Wall Street Journal*, 13 de abril, 1999). Incluso en medio de todas estas fuertes críticas, las encuestas por Internet se están volviendo cada vez más comunes. ¿Qué preocupaciones, si las hay, le despiertan las encuestas por Internet?

**7.57** Un estudio realizado por Rajesh Mirani y Albert Lederer ("An Instrument for Accessing the Organizational Benefits of IS Projects", *Decision Sciences*, vol. 29, 1998, 803-838) analiza los beneficios organizacionales de los proyectos de sistemas de información (IS, por sus siglas en inglés). Los investigadores enviaron por correo 936 cuestionarios, dirigidos a los miembros, seleccionados al azar, de una enorme organización de sistemas de información que abarca a todo el país. Recibieron 200 respuestas válidas, para una tasa de respuesta del 21%. De los 200 encuestados, 190 respondieron preguntas referentes a un proyecto de IS recién terminado. El presupuesto promedio para estos proyectos fue de \$3.8 millones, con un rango que oscila entre 4,000 y 100 millones de dólares. De esas 190 respuestas, el 45% señalaban que se solicitó la aprobación del director general antes de iniciar los proyectos.

- ¿Cuál fue la fuente de datos utilizada en este estudio?
- Analice el método de muestreo utilizado en este estudio.
- ¿Qué tipos de errores de encuesta cree que es más probable que encuentren los investigadores?

**7.58** Como parte del proceso de mediación supervisado por un juez federal, cuyo objeto era dar fin a un juicio en el que se acusa a Cincinnati, Ohio, de décadas de discriminación contra los afroamericanos, se realizaron encuestas sobre cómo mejorar las relaciones entre la policía de Cincinnati y la comunidad. Se envió una encuesta a los 1,020 miembros del cuerpo de policía de Cincinnati. La encuesta incluyó una carta en la que el jefe de policía y el presidente de la Fraternidad de Policías alentaban la participación. Los participantes podían devolver sus respuestas por escrito o responder la encuesta en línea. Para consternación de los investigadores, sólo se completaron 158 encuestas ("Few Cops Fill Out Survey", *The Cincinnati Enquirer*, 22 de agosto, 2001, B3).

- ¿Qué tipo de errores o sesgos deben preocupar de manera especial a los investigadores?
- ¿Qué paso(s) deben emprender los investigadores para tratar de compensar los problemas observados en el inciso a)?
- ¿Qué se pudo efectuar de diferente manera para mejorar las virtudes de la encuesta?

**7.59** De acuerdo con la encuesta realizada por International Communications Research para Capital One Financial, el 24% de los adolescentes de entre 13 y 19 años de edad tienen un teléfono celular y el 10% de ellos cuentan con localizador ("USA Snapshots", *USA Today*, 16 de agosto, 2001, A1).

- ¿Qué otra información le gustaría conocer antes de aceptar los resultados de esta encuesta?
- Suponga que desea realizar una encuesta semejante en la región geográfica donde vive. Describa la población para su encuesta.
- Explique cómo podría reducir al mínimo la posibilidad de un error de cobertura en este tipo de encuesta.

- d. Explique cómo podría reducir al mínimo la posibilidad de un error de no respuesta en este tipo de encuesta.
- e. Explique cómo podría reducir al mínimo la posibilidad de un error de muestreo en este tipo de encuesta.
- f. Explique cómo podría reducir al mínimo la posibilidad de un error de medición en este tipo de encuesta.

**7.60** De acuerdo con la doctora Sarah Beth Estes, profesora de sociología en la Universidad de Cincinnati, y con la doctora Jennifer Glass, profesora de sociología en la Universidad de Iowa, las mujeres trabajadoras que aprovechan los horarios ventajosos por motivos familiares pueden obtener menores salarios. De manera más específica, la sociólogas reportan que en un estudio realizado con 300 mujeres que trabajan y tienen hijos, y vuelven a laborar aprovechando programas de flexibilidad o aquellos que permiten realizar el trabajo en casa, se observó que obtenían aumentos de sueldo entre un 16 y un 26% menores, en promedio, que otros trabajadores (“Study: ‘Face Time’ Can Affect Moms’ Raises”, *The Cincinnati Enquirer*, 28 de agosto, 2001, A1).

- a. ¿Qué otra información le gustaría conocer antes de aceptar los resultados de esta encuesta?
- b. Si usted realizará un estudio similar en el área geográfica donde vive, defina la población, marco y método de muestreo que podría utilizar.

**7.61 (Proyecto de grupo)** La tabla de números aleatorios es un ejemplo de distribución uniforme, porque cada uno de sus dígitos tiene igual posibilidad de presentarse. Comenzando en la fila correspondiente al día del mes en el que nació, utilice la tabla de números aleatorios (tabla E.1) para elegir un dígito a la vez.

Seleccione cinco muestras distintas de  $n = 2$ ,  $n = 5$  y  $n = 10$ . Calcule la media muestral de cada una de ellas. Desarrolle una distribución de frecuencia de las medias muestrales para los resultados de toda la clase, con base en muestras de tamaños  $n = 2$ ,  $n = 5$  y  $n = 10$ .

¿Qué se concluye respecto a la forma de la distribución muestral de cada uno de estos tamaños de la muestra?

**7.62 (Proyecto de grupo)** Lance una moneda 10 veces y registre el número de caras. Si cada uno de los alumnos realiza cinco veces este experimento, se puede desarrollar una distribución de frecuencia del número de caras a partir de los resultados de todo el salón. ¿Esta distribución parece aproximarse a la distribución normal?

**7.63 (Proyecto de grupo)** El número de automóviles en espera, formados en un servicio de lavado, se distribuye de la siguiente manera:

Número de automóviles	Probabilidad
0	0.25
1	0.40
2	0.20
3	0.10
4	0.04
5	0.01

Usted utiliza la tabla de números aleatorios (tabla E.1) para seleccionar muestras a partir de esta distribución, asignando números para la fila de espera, como se expresa continuación:

1. Comenzando en la fila correspondiente al día del mes en el que usted nació.
2. Seleccionando un número aleatorio de dos dígitos.
3. Si selecciona un número aleatorio de 00 a 24, registre una longitud de 0; si es de 25 a 64, registre una longitud de 1; si es de 65 a 84, registre una longitud de 2; si es de 85 a 94, registre una longitud de 3; si es de 95 a 98, registre una longitud de 4; si es de 99, registre una longitud de 5.

Seleccione muestras de  $n = 2$ ,  $n = 5$  y  $n = 10$ . Calcule la media de cada muestra. Por ejemplo, si la muestra con un tamaño de 2 tiene como resultado los números aleatorios 18 y 46, estos corresponderán a las longitudes de la fila de espera de 0 y 1, respectivamente, produciendo una media muestral de 0.5. Si cada estudiante selecciona cinco muestras diferentes para cada tamaño de la muestra, se puede desarrollar una frecuencia de distribución de las medias muestrales (para cada tamaño de la muestra) a partir de los resultados de todo el salón. ¿Qué conclusiones se obtienen con respecto a la distribución muestral de la media conforme aumenta el tamaño de la muestra?

**7.64 (Proyecto de grupo)** La tabla de números aleatorios permite simular la selección de distintas pelotas de colores en un recipiente, de la siguiente manera:

1. Comenzando en la fila correspondiente al día del mes en el que usted nació.
2. Seleccione números de un dígito.
3. Si selecciona un dígito aleatorio entre 0 y 6, considere que la pelota es blanca; si el dígito aleatorio es 7, 8 o 9, considere que la pelota es roja.

Seleccione muestras de  $n = 10$ ,  $n = 25$  y  $n = 50$  dígitos. En cada una de las muestras, cuente el número de pelotas blancas y calcule su proporción en la muestra. Si cada estudiante de la clase selecciona cinco muestras diferentes para cada tamaño de la muestra, puede desarrollar una distribución de frecuencia para la proporción de bolas blancas (para cada tamaño de la muestra) a partir de los resultados de todo el salón. ¿Qué conclusiones se obtienen con respecto a la distribución muestral de la proporción conforme aumenta el tamaño de la muestra?

**7.65 (Proyecto de grupo)** Suponga que el paso 3 del problema 7.64 utiliza la siguiente regla: “Si selecciona un dígito aleatorio entre 0 y 8, considere que la pelota es blanca; si el dígito aleatorio es 9, considere que la pelota es roja”. Compare y establezca las diferencias entre los resultados de este problema y los del problema 7.64.

## CASO ACTUAL

### ADMINISTRACIÓN DEL SPRINGVILLE HERALD

Para continuar el mejoramiento de la calidad en el caso que describe el capítulo 6 “Administración del Springville Herald”, el departamento de producción ha estado supervisando la tonalidad oscura de la impresión del periódico. Como antes, la tonalidad oscura se mide en una escala estándar donde el valor objetivo es 1.0. Los datos recabados durante el año anterior señalan que la tonalidad oscura tiene una distribución normal, con una media de 1.005 y una desviación estándar de 0.10.

SH7.1 Cada día, se seleccionan 25 puntos del primer ejemplar y se mide su tonalidad oscura. Suponiendo que la distribución no ha cambiado con respecto a la del año

pasado, ¿cuál es la probabilidad de que la tonalidad oscura media de los puntos sea:

- menor que 1.0?
- entre 0.95 y 1.0?
- entre 1.0 y 1.05?
- menor que 0.95 o mayor que 1.05?
- Suponga que la media de la muestra de 25 puntos tomada el día de hoy es 0.952. Con base en este resultado, ¿qué conclusión se obtiene con respecto a la tonalidad oscura del periódico? Explique su respuesta.

## CASO WEB

*Aplique sus conocimientos sobre las distribuciones muestrales a este caso Web, que reconsidera el escenario “Uso de la estadística” de la Oxford Cereals.*

La organización TriCities Consumers Concerned About Cereal Companies That Cheat (TCCACCT) sospecha que las empresas de cereales, incluida la Oxford Cereals, engaña a los consumidores al empacar los cereales con menos peso del señalado en la etiqueta. Visite el sitio Web de la organización en [www.prenhall.com/Springville/CerealCheaters.htm](http://www.prenhall.com/Springville/CerealCheaters.htm), examine sus quejas y datos de respaldo, y luego responda lo siguiente:

1. ¿Yerran los procedimientos de recopilación de datos utilizados por la TCCACCT para obtener sus conclusiones? ¿Cuáles procedimientos podría seguir el grupo para hacer más rigurosos sus análisis?
2. Suponga que las dos muestras compuestas por cinco cajas de cereal (una muestra por cada una de las dos variedades) mencionadas en el sitio Web de la TCCACCT se recogieron de manera aleatoria. Para cada una de las muestras, haga lo siguiente:

- a. Calcule la media muestral.
- b. Suponga que la desviación estándar del proceso es 15 gramos. Calcule el porcentaje de todas las muestras de cada uno de los procesos que tendría una media muestral menor que el valor calculado en el inciso a).
- c. Una vez más, suponiendo que la desviación estándar es de 15 gramos, calcule el porcentaje de cajas de cereal individuales que tendrán un peso inferior al valor calculado del inciso a).
3. ¿Qué conclusiones, si las hay, se obtienen al utilizar sus cálculos sobre los procesos de llenado de dos cereales distintos?
4. Un representante de la Oxford Cereals solicitó que la TCCACCT elimine su página donde analiza los faltantes en las cajas de Oxford Cereals. ¿Es razonable esta petición? ¿Por qué?
5. ¿Se pueden emplear las técnicas incluidas en este capítulo como lo hace la TCCACCT para demostrar el engaño? ¿Por qué?

## REFERENCIAS

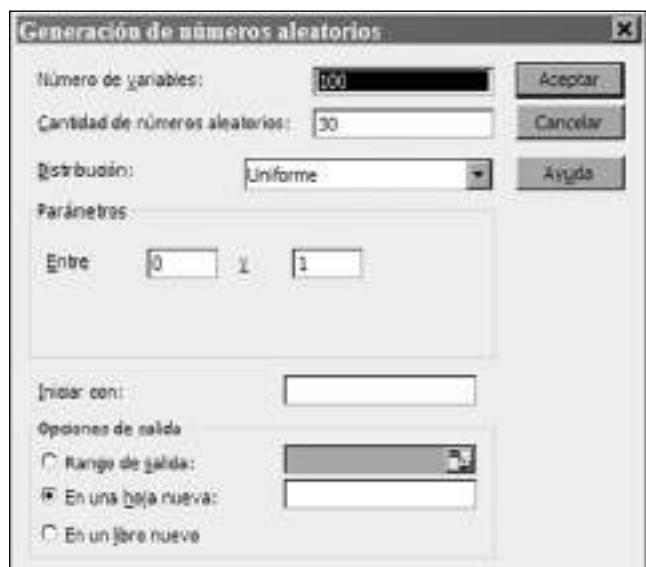
1. Cochran, W. G., *Sampling Techniques*, 3a. ed. (Nueva York: Wiley, 1977).
2. Gallup, G. H., *The Sophisticated Poll-Watcher's Guide* (Princeton, NJ: Princeton Opinion Press, 1972).
3. Goleman, D., “Pollsters Enlist Psychologists in Quest for Unbiased Results”, *The New York Times*, 7 de septiembre, 1993, C1 y C11.
4. Levine, D. M., P. Ramsey y R. Smidt, *Applied Statistics for Engineers and Scientists Using Microsoft Excel and Minitab* (Upper Saddle River, NJ: Prentice Hall, 2001).
5. Microsoft Excel 2003 (Redmond, WA: Microsoft Corp., 2003).
6. Minitab for Windows Version 14 (State College, PA: Minitab Inc., 2004).
7. Mosteller, F. et al., *The Pre-Election Polls of 1948* (Nueva York: Social Science Research Council, 1949).
8. Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Nueva York: The Free Press, 1955).

# Apéndice 7 Uso de software para las distribuciones muestrales

## A7.1 EXCEL

### Para la simulación de la distribución muestral

Use el ToolPak para análisis de datos. Abra la hoja de trabajo que contiene los datos para los que desea generar una simulación de distribución muestral. Seleccione **Herramientas** → **Análisis de datos**. En la lista que aparece en la ventana de diálogo Análisis de datos, seleccione **Generación de número aleatorio** y dé clic en **OK**. En el cuadro de diálogo Generador de Número Aleatorio (vea la figura A7.1), introduzca el número de muestras como el **Número de variables** y el tamaño de la muestra de cada muestra como la **Cantidad de números aleatorios**. Seleccione el tipo de distribución en el menú desplegable **Distribución** y realice las entradas necesarias en el área **Parámetros** (los parámetros varían de acuerdo con la distribución seleccionada). Seleccione el botón con la opción **Crear nueva hoja** y dé clic en el botón **OK**. En la figura A7.1 se muestran las entradas para generar 100 muestras de  $n = 30$  a partir de una población con distribución uniforme.



**FIGURA A7.1** Ventana de diálogo generador de números aleatorios para análisis de datos.

**O** Vea la sección G.10 (**Simulación de distribución muestral**) si desea que PHStat2 genere una versión mejorada de esta hoja.

## A7.2 MINITAB

Para desarrollar una simulación de la distribución muestral de la media a partir de una población con distribución uniforme con 100 muestras de  $n = 30$ , seleccione **Calc** → **Random Data** → **Uniform**. En la ventana de diálogo Uniform Distribution (vea la figura A7.2):

1. Introduzca **100** en el cuadro Generate rows of data (Generar filas de datos).
2. Introduzca **C1-C30** en el cuadro de edición Store in column(s).
3. Introduzca **0.0** en el cuadro de edición Lower endpoint y **1.0** en el Upper endpoint. Dé clic en el botón **OK**.

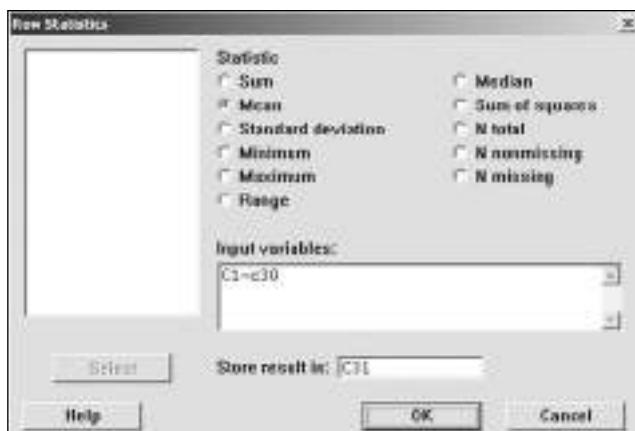


**FIGURA A7.2** Ventana de diálogo Uniform Distribution de Minitab.

Ahora se introducen 100 filas de valores en las columnas C1-C30. Para calcular los estadísticos de fila para cada una de las 100 muestras, seleccione **Calc** → **Row Statistics**, y luego, en la ventana de diálogo Row Statistics (vea la figura A7.3):

1. Seleccione el botón **Mean**.
2. Introduzca **C1-C30** en el cuadro de edición Input variables. Introduzca **C31** en el cuadro de edición Store result in. Dé clic en el botón **OK**.

La media de cada una de las 100 muestras se guarda en la columna C31. Para calcular los estadísticos del conjunto de 100 medias muestrales, seleccione **Stat** → **Basic Statistics** → **Display Descriptive Statistics**. Introduzca **C31** en el cuadro de edición Variables. Dé clic en el botón **OK**.



**FIGURA A7.3** Ventana de diálogo Row Statistics de Minitab.

Para generar un histograma de las 100 medias muestrales, seleccione **Graph → Histogram** y haga lo siguiente:

1. En la ventana de diálogo Histograms, seleccione **Simple**. Dé clic en el botón **OK**.
2. En la ventana de diálogo Histogram-Simple, introduzca **C31** en el cuadro Graph variables. Dé clic en el botón **OK**.

Para desarrollar una simulación de la distribución muestral de la media para una población normal, seleccione **Calc → Random Data → Normal**. Introduzca el valor de  $\sigma$  en el cuadro Mean y para  $\mu$  en el cuadro de edición Standard deviation. Siga el resto de las instrucciones indicadas para la población uniforme.



# CAPÍTULO 8

## Estimación de intervalos de confianza

---

**USO DE LA ESTADÍSTICA:** Factura de ventas auditadas en la empresa de remodelaciones Saxon

- 8.1 ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$  CONOCIDA)**
- 8.2 ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$  DESCONOCIDA)**
  - Distribución  $t$  Student
  - Propiedades de la distribución  $t$
  - El concepto de grados de libertad
  - El establecimiento del intervalo de confianza
- 8.3 ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN**
- 8.4 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA**

- Determinación del tamaño de la muestra para la media
- Determinación del tamaño de la muestra para la proporción
- 8.5 ESTIMACIÓN DEL INTERVALO DE CONFIANZA Y CONSIDERACIONES ÉTICAS**
- A.8 USO DEL SOFTWARE PARA LOS INTERVALOS DE CONFIANZA Y LA DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA**
  - A8.1 Excel*
  - A8.2 Minitab*
  - A8.3 (Tema del CD-ROM) SPSS*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- A construir e interpretar estimaciones de intervalo de confianza para la media y la proporción
- Cómo determinar el tamaño de la muestra necesario para desarrollar un intervalo de confianza para la media o la proporción

# USO DE LA ESTADÍSTICA



## Facturas de ventas auditadas en la empresa de remodelaciones Saxon

La empresa de remodelaciones Saxon distribuye productos para el mantenimiento del hogar en el noreste de Estados Unidos. Como contador de la empresa, usted es responsable de la exactitud en la administración de un inventario integrado y del sistema de información de ventas. Podría revisar los contenidos de cada registro para verificar la exactitud del sistema, pero una revisión muy detallada resulta costosa, además de que requiere mucho tiempo. Un mejor método es usar las técnicas de estadística inferencial para obtener conclusiones acerca de la población de todos los registros a partir de una muestra relativamente pequeña recolectada durante una auditoría. Al final de cada mes, selecciona una muestra de las facturas de ventas para determinar lo siguiente:

- La media de las cantidades de dólares indicadas en las facturas de ventas.
- La proporción de facturas de ventas que contienen errores.

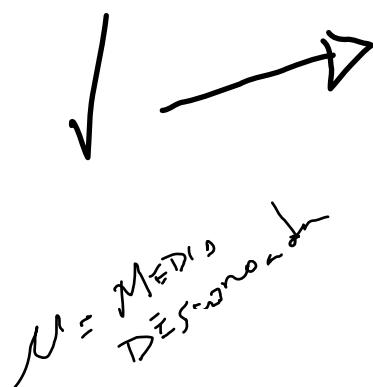
¿Qué tan precisos son estos resultados extraídos a partir de las muestras y cómo usar esta información? ¿Son las muestras lo suficientemente grandes para aportar la información que necesita?

**L**a estadística inferencial es el proceso de uso de los resultados derivados de las muestras para obtener conclusiones acerca de las características de una población. La estadística inferencial nos permite *estimar* características desconocidas como la media de la población o la proporción de la población. Existen dos tipos de estimaciones usadas para estimar los parámetros de la población: la estimación puntual y la estimación de intervalo. Una **estimación puntual** es el valor de un solo estadístico de muestra. Una **estimación del intervalo de confianza** es un rango de números, llamado intervalo, construido alrededor de la estimación puntual. El intervalo de confianza se construye de manera que la probabilidad del parámetro de la población se localice en algún lugar dentro del intervalo conocido.

Suponga que quiere estimar la media del puntaje GPA (Grade Point Average) de todos los alumnos en su universidad. La media del GPA para todos los alumnos es una media desconocida de la población, simbolizada como  $\mu$ . Usted selecciona una muestra de alumnos y encuentra que la media es de 2.80. La muestra de la media  $\bar{X} = 2.80$  es la estimación puntual de la media poblacional  $\mu$ . ¿Qué tan preciso es el 2.80? Para responder esta pregunta debe construir una estimación del intervalo de confianza.

En este capítulo aprenderá cómo construir e interpretar estimaciones de intervalo de confianza. Recuerde que la media de la muestra  $\bar{X}$  es una estimación puntual de la media poblacional  $\mu$ . Sin embargo, la media de la muestra puede variar de una muestra a otra porque depende de los elementos seleccionados en la muestra. Tomando en cuenta la variabilidad de muestra a muestra (vea la sección 7.2 sobre la distribución muestral de la media), aprenderá a desarrollar la estimación del intervalo para la media poblacional. El intervalo construido tendrá una confianza especificada de la estimación correcta del valor del parámetro poblacional  $\mu$ . En otras palabras, existe una confianza específica de que  $\mu$  se encuentre en algún lugar en el rango de números definidos por el intervalo.

Suponga que después de estudiar este capítulo, encuentra que un intervalo de confianza del 95% para la media del GPA de su universidad es  $(2.75 \leq \mu \leq 2.85)$ . Puede interpretar esta estimación del intervalo estableciendo que está 95% confiado en que la media del GPA en su universidad esté entre 2.75 y 2.85. Hay un 5% de posibilidades de que la media del GPA esté por debajo de 2.75 o por arriba de 2.85.



Después de aprender acerca del intervalo de confianza para la media, aprenderá cómo desarrollar un intervalo de estimación para la proporción de la población. Entonces comprenderá qué tan grande debe ser la muestra seleccionada al construir intervalos de confianza.

## 8.1 ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$ CONOCIDA)

En la sección 7.2 usó el teorema de límite central y el conocimiento de la distribución poblacional para determinar el porcentaje de las medias de la muestra que caen dentro de ciertas distancias de la media poblacional. Por ejemplo, para el caso del llenado de cereal descrito en el capítulo 7 (vea la página 213), el 95% de las medias muestrales están entre 362.12 y 373.88 gramos. Esta afirmación se basa en el *razonamiento deductivo*. Sin embargo, lo que usted necesita aquí es el *razonamiento inductivo*.

Se requiere del razonamiento inductivo porque en la estadística inferencial se usan los resultados de una sola muestra para obtener conclusiones sobre la población, y no a la inversa. Suponga que en el ejemplo del llenado del cereal, usted quisiera estimar la media poblacional desconocida usando la información de una sola muestra. Por lo tanto, en vez de tomar  $\mu \pm (1.96)(\sigma/\sqrt{n})$  para encontrar los límites inferior y superior alrededor de la  $\mu$  como en la sección 7.2, se sustituye la media de la muestra  $\bar{X}$  por la  $\mu$  desconocida y se utiliza  $\bar{X} \pm (1.96)(\sigma/\sqrt{n})$  como un intervalo para estimar la  $\mu$  desconocida. Aunque en la práctica se selecciona una sola muestra de tamaño  $n$  y se calcula la media  $\bar{X}$ , para comprender el significado total de la estimación de intervalo, usted requiere examinar un conjunto hipotético de todas las posibles muestras de valores  $n$ .

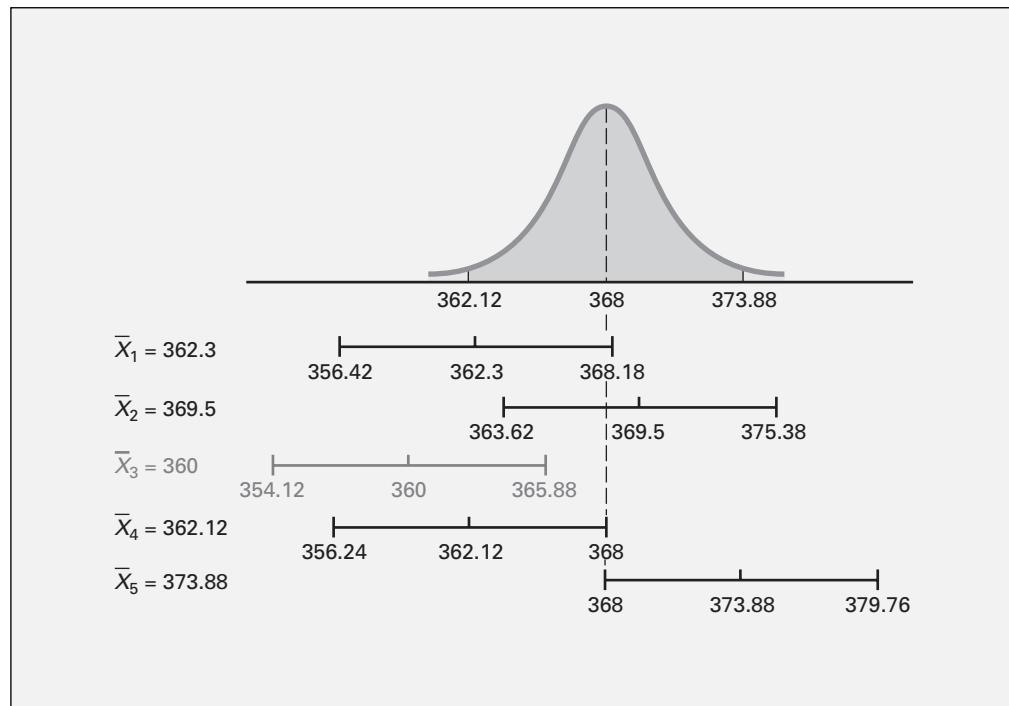
Suponga que la muestra de  $n = 25$  cajas tiene una media de 362.3 gramos. El intervalo desarrollado para estimar  $\mu$  es  $362.3 \pm (1.96)(15)/(\sqrt{25})$  o  $362.3 \pm 5.88$ . El estimado de  $\mu$  es

$$356.42 \leq \mu \leq 368.18$$

Como la media poblacional  $\mu$  (igual a 368) está incluida dentro del intervalo, esta muestra ha llevado a una correcta afirmación acerca de  $\mu$  (vea la figura 8.1).

**FIGURA 8.1**

Estimación del intervalo de confianza para cinco diferentes muestras de  $n = 25$  tomadas de una población donde  $\mu = 368$  y  $\sigma = 15$ .



Para continuar con este ejemplo hipotético, suponga que para una muestra diferente de  $n = 25$  cajas, la media es 369.5. El intervalo desarrollado para esta muestra es

$$369.5 \pm (1.96)(15)/(\sqrt{25})$$

o  $369.5 \pm 5.88$ . El estimado es

$$363.62 \leq \mu \leq 375.38$$

Puesto que la media poblacional  $\mu$  (igual a 368) está incluida también dentro de este intervalo, esta afirmación sobre  $\mu$  es correcta.

Ahora bien, antes de que piense que siempre que se desarrolla una estimación del intervalo de confianza se hacen afirmaciones correctas sobre  $\mu$ , suponga que una tercera hipótesis de una muestra de  $n = 25$  cajas es seleccionada y la media de la muestra es igual a 360 gramos. El intervalo desarrollado aquí es  $360 \pm (1.96)(15)/(\sqrt{25})$  o  $360 = \pm 5.88$ . En este caso, la estimación de  $\mu$  es

$$354.12 \leq \mu \leq 365.88$$

Esta estimación *no* es correcta, porque la media poblacional no está incluida en el intervalo desarrollado para esta muestra (vea la figura 8.1 en la página 239). Así, para algunas muestras, la estimación del intervalo de  $\mu$  es correcta, pero para otras es incorrecta. En la práctica sólo se selecciona una muestra, y puesto que la media poblacional es desconocida, no es posible determinar si la estimación del intervalo es correcta.

Para resolver el dilema de tener en ocasiones un intervalo que proporcione una estimación correcta y en otras tener un intervalo que proporcione una estimación incorrecta, se requiere determinar la proporción de muestras que generan intervalos que dan por resultado afirmaciones correctas acerca de la media poblacional  $\mu$ . Para hacerlo, considere dos diferentes muestras hipotéticas: el caso en el que  $\bar{X} = 362.12$  gramos y el caso en que  $\bar{X} = 373.88$  gramos. Si  $\bar{X} = 362.12$ , el intervalo es  $362.12 \pm (1.96)(15)/(\sqrt{25})$  o  $362.12 \pm 5.88$ . Esto lleva al siguiente intervalo

$$356.24 \leq \mu \leq 368.00$$

Como la media poblacional de 368 está en el límite superior del intervalo, la afirmación es correcta (vea la figura 8.1).

Cuando  $\bar{X} = 373.88$ , el intervalo es  $373.88 \pm (1.96)(15)/(\sqrt{25})$  o  $373.88 \pm 5.88$ . El intervalo para la media de la muestra es

$$368.00 \leq \mu \leq 379.76$$

En este caso, puesto que la media poblacional de 368 está incluida en el límite inferior del intervalo, la afirmación es correcta.

En la figura 8.1, se observa que cuando la media de la muestra cae en cualquier lugar entre 362.12 y 373.88 gramos, la media poblacional está incluida en *algún lugar* dentro del intervalo. En la sección 7.2 en la página 213, usted encontró que el 95% de las medias de la muestra caen entre 362.12 y 373.88 gramos. Por lo tanto, el 95% de todas las muestras de  $n = 25$  cajas tienen medias de muestra que incluyen la media poblacional dentro del intervalo desarrollado. Nos referimos al intervalo de 362.12 a 373.88 como intervalo con un 95% de confianza.

Como en la práctica usted sólo elige una muestra y la  $\mu$  es desconocida, nunca sabe con certeza si el intervalo específico incluye la media poblacional o no. Sin embargo, si toma todas las posibles muestras de  $n$  y calcula sus medias de muestra, el 95% de todos los intervalos incluirán la media poblacional, y sólo el 5% de ellas no. En otras palabras, tiene un 95% de confianza de que la media poblacional se encuentra en algún lugar dentro del intervalo. Así, podrá interpretar el anterior intervalo de confianza de la siguiente manera:

“Estoy 95% confiado en que la media de la cantidad de cereal en la población de cajas se encuentra entre 362.12 y 373.88 gramos.”

En algunas situaciones, usted quisiera tener un mayor grado de confianza (como el 99%) de incluir la media poblacional dentro del intervalo. En otros casos, podría aceptar un menor grado de confianza (como el 90%), al estimar la media poblacional de forma correcta.

En general, el **nivel de confianza** se simboliza con  $(1 - \alpha) \times 100\%$ , donde  $\alpha$  es la proporción de las colas de la distribución que están fuera del intervalo de confianza. La proporción de la cola superior de la distribución es  $\alpha/2$ , y la proporción de la cola inferior de la distribución es  $\alpha/2$ . Utilice la ecuación (8.1) para construir una estimación de intervalo de confianza  $(1 - \alpha) \times 100\%$  de la media con una  $\sigma$  conocida.

### INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$ CONOCIDA)

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

o

$$\bar{X} - Z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z \frac{\sigma}{\sqrt{n}} \quad (8.1)$$

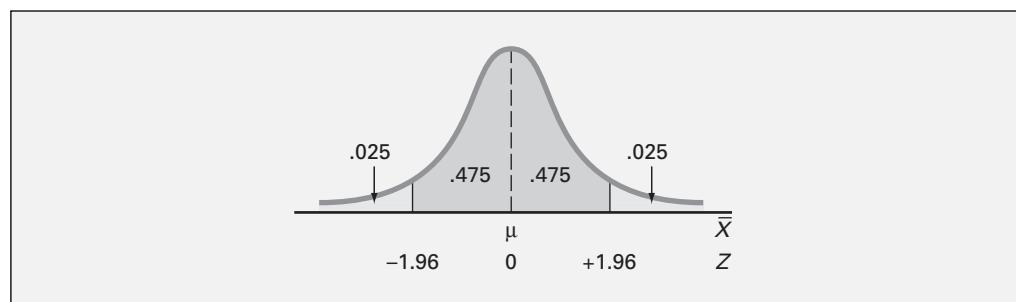
donde  $Z$  = valor correspondiente a un área acumulativa de  $1 - \alpha/2$  de la distribución normal estandarizada, esto es, una probabilidad de la cola superior de  $\alpha/2$ .

Se llama **valor crítico** al valor de  $Z$  necesario para construir un intervalo de confianza para la distribución. El 95% de confianza corresponde a un valor  $\alpha$  de 0.05. El valor crítico  $Z$  correspondiente al área acumulativa de 0.9750 es 1.96 porque hay 0.025 en la cola superior de la distribución y el área acumulativa menor a  $Z = 1.96$  es 0.975.

Hay un valor crítico diferente para cada nivel de confianza  $1 - \alpha$ . Un nivel de confianza del 95% lleva a un valor  $Z$  de 1.96 (vea la figura 8.2). El 99% de confianza corresponde a un valor  $\alpha$  de 0.01. El valor de  $Z$  es aproximadamente 2.58 porque el área de la cola alta es 0.005 y el área acumulativa menor a  $Z = 2.58$  es 0.995 (vea la figura 8.3).

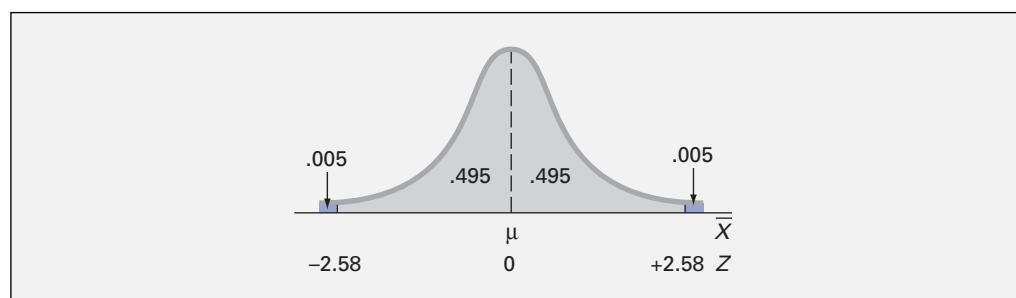
**FIGURA 8.2**

Curva normal para determinar el valor de  $Z$  necesario para el 95% de confianza.



**FIGURA 8.3**

Curva normal para determinar el valor  $Z$  necesario para el 99% de confianza.



$Z =$   
 1.96  
 2.58

Ahora que hemos considerado diferentes niveles de confianza, ¿por qué no llevar el intervalo de confianza lo más cerca posible del 100%? Antes de hacerlo, usted necesita darse cuenta de que cualquier incremento en el nivel de confianza se logra únicamente extendiendo (y haciendo menos preciso) el intervalo de confianza. Aquí no hay ningún “almuerzo gratis”. Tendría más confianza de que la media poblacional esté dentro de un rango de valores más amplio. Sin embargo, esto podría llevar a la interpretación de que el intervalo de confianza es menos útil. El intercambio entre la amplitud del intervalo de confianza y el nivel de confianza se discutirá con mayor profundidad dentro del contexto de determinación del tamaño de la muestra en la sección 8.4.

El ejemplo 8.1 ilustra la aplicación de la estimación del intervalo de confianza.

### EJEMPLO 8.1

#### ESTIMACIÓN DE LA MEDIA DE LA LONGITUD DEL PAPEL CON UN 95% DE CONFIANZA

Un fabricante de papel para computadora tiene un proceso de producción que opera continuamente a lo largo del turno completo. Se espera que el papel tenga una media de longitud de 11 pulgadas y una desviación estándar de 0.02 pulgadas. A intervalos periódicos, se selecciona una muestra para determinar si la media de longitud del papel es igual a 11 pulgadas o para ver si algo ha salido mal durante el proceso que haya cambiado la longitud del papel que se fabrica. Usted selecciona una muestra aleatoria de 100 hojas; la media de longitud del papel es de 10.998 pulgadas. Construya una estimación del intervalo de confianza del 95% para la media poblacional de la longitud del papel.

**SOLUCIÓN** Se emplea la ecuación (8.1) de la página 241, con  $Z = 1.96$  para un 95% de confianza,

$$\begin{aligned}\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} &= 10.998 \pm (1.96) \frac{0.02}{\sqrt{100}} \\ &= 10.998 \pm 0.00392 \\ 10.99408 \leq \mu &\leq 11.00192\end{aligned}$$

Así, con un 95% de confianza, usted concluye que la media poblacional está entre 10.99408 y 11.00192 pulgadas. Como el intervalo incluye al 11, valor que indica que el proceso de producción funciona adecuadamente, no hay razón para creer que algo está mal con el proceso de fabricación. Para ver el efecto de usar un intervalo de confianza del 99%, examine el ejemplo 8.2.

Para ver el efecto de usar un intervalo de confianza del 99%, examine el ejemplo 8.2.

### EJEMPLO 8.2

#### ESTIMACIÓN DE LA MEDIA DE LA LONGITUD DEL PAPEL CON UN 99% DE CONFIANZA

Construya una estimación de intervalo de confianza del 99% para la media poblacional de la longitud del papel.

**SOLUCIÓN** Usando la ecuación (8.1) de la página 241, con  $Z = 2.58$  para un nivel de confianza del 99%,

$$\begin{aligned}\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} &= 10.998 \pm (2.58) \frac{0.02}{\sqrt{100}} \\ &= 10.998 \pm 0.00516 \\ 10.99284 \leq \mu &\leq 11.00316\end{aligned}$$

Una vez más, puesto que el 11 está incluido dentro de la amplitud del intervalo, no hay razón para creer que algo está mal con el proceso de producción.

## PROBLEMAS PARA LA SECCIÓN 8.1

### Aprendizaje básico

ASISTENCIA  
de PH Grade

- 8.1** Si  $\bar{X} = 85$ ,  $\sigma = 8$  y  $n = 64$ , construya una estimación de intervalo de confianza del 95% de la media poblacional de  $\mu$ .

ASISTENCIA  
de PH Grade

- 8.2** Si  $\bar{X} = 125$ ,  $\sigma = 24$  y  $n = 36$ , construya una estimación de intervalo de confianza del 99% para la media poblacional de  $\mu$ .

- 8.3** Un investigador de mercado afirma que tiene un nivel de confianza del 95% en que la media de las ventas mensuales de un producto está entre \$170,000 y \$200,000. Explique el significado de su afirmación.

- 8.4** ¿Por qué no es posible tener un 100% de confianza en el ejemplo 8.1 de la página 242?

- 8.5** De los resultados del ejemplo 8.1, en la página 242, en relación con la producción de papel, ¿es verdad que el 95% de las medias de la muestra caerán entre 10.99408 y 11.00192 pulgadas? Explique su respuesta.

- 8.6** En el ejemplo 8.1 de la página 242, ¿es cierto que no puede estar seguro si la media poblacional está entre 10.99408 y 11.00192 pulgadas? Explique su respuesta.

### Aplicación de conceptos

ASISTENCIA  
de PH Grade

- 8.7** El gerente de una tienda de artículos para pintar quiere estimar la cantidad real de pintura contenida en latas de 1 galón compradas con las de un productor conocido a nivel nacional. Se sabe, a partir de las especificaciones del fabricante, que la desviación estándar de la cantidad de pintura es igual a 0.02 galón. Se selecciona una muestra aleatoria de 50 latas y la media muestral de la cantidad de pintura contenida en una lata es de 0.995 por galón.

- Construya una estimación de intervalo de confianza del 99% de la media poblacional cantidad de pintura contenida en una lata de 1 galón.
- Con base en sus resultados, ¿cree usted que el gerente tiene el derecho de quejarse con el fabricante? ¿Por qué?
- ¿Debe suponerse que la población cantidad de pintura por lata se distribuye normalmente? Explique su respuesta.

- d.** Construya una estimación de intervalo de confianza del 95%. ¿Cómo cambia esto su respuesta al inciso b)?

ASISTENCIA  
de PH Grade

AUTO  
Examen

- 8.8** El gerente de control de calidad de una fábrica de focos necesita estimar la media de vida de un gran embarque de focos. La desviación estándar es de 100 horas. Una muestra aleatoria de 64 focos indicó que la vida media de la muestra es de 350 horas.

- Construya una estimación de intervalo de confianza del 95% para la media poblacional de vida de los focos de este embarque.
- ¿Cree que el fabricante tiene el derecho de afirmar que los focos tienen un promedio de vida de 400 horas? Explique por qué.
- ¿Debe suponerse que la vida de la población de focos se distribuye normalmente? Explique por qué.
- Suponga que la desviación estándar cambió a 80 horas. ¿Cuáles son sus respuestas para los incisos a) y b)?

ASISTENCIA  
de PH Grade

- 8.9** La división de inspección de Pesos y Medidas del condado de Lee desea estimar la cantidad real de contenido en botellas de 2 litros de bebida refrescante en la planta embotelladora local de una empresa conocida a nivel nacional. La planta embotelladora ha informado a la división de inspección que la desviación estándar poblacional para las botellas de 2 litros es de 0.05 litro. Una muestra aleatoria de 100 botellas de 2 litros en la planta embotelladora indica una media muestral de 1.99 litros.

- Construya una estimación de intervalo de confianza del 95% de la media poblacional cantidad de bebida refrescante en cada botella.
- ¿Debe suponerse que la población de relleno de refresco se distribuye normalmente? Explique por qué.
- Explique por qué el valor de 2.02 litros para una botella sola no es inusual, aun cuando esté fuera del intervalo de confianza calculado.
- Suponga que la media muestral hubiera sido de 1.97 litros. ¿Cuál sería su respuesta al inciso a)?

## 8.2 ESTIMACIÓN DE INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$ DESCONOCIDA)

Así como la media poblacional  $\mu$  suele ser desconocida, rara vez se conoce la desviación estándar real de la población  $\sigma$ . Por lo tanto, usted requiere desarrollar una estimación del intervalo de confianza de  $\mu$  usando sólo los estadísticos de muestra  $\bar{X}$  y  $S$ .

### Distribución t Student

Al comenzar el siglo XX, un especialista en estadística de la Guinness Breweries en Irlanda (vea la referencia 3) llamado William S. Gosset deseaba hacer inferencias acerca de la media cuando la  $\sigma$  fuera desconocida. Como a los empleados de Guinness no se les permitía publicar el trabajo de in-

vestigación bajo sus propios nombres, Gosset adoptó el seudónimo de “Student”. La distribución que desarrolló se conoce como la **distribución *t* Student**.

Si la variable aleatoria  $X$  se distribuye normalmente, entonces el siguiente estadístico tiene una distribución  $t$  con  $n - 1$  **grados de libertad**.

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

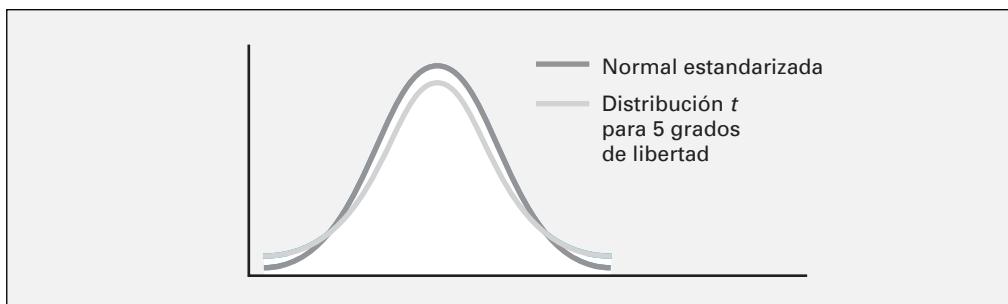
Esta expresión tiene la misma forma que el estadístico  $Z$  en la ecuación (7.4) de la página 211, con la excepción de que  $S$  se usa para estimar la  $\sigma$  desconocida. El concepto de *grados de libertad* se explica más adelante en la página 245.

## Propiedades de la distribución $t$

En apariencia, la distribución  $t$  es muy similar a la distribución normal estandarizada. Ambas distribuciones tienen forma de campana. Sin embargo, la distribución  $t$  tiene mayor área en los extremos y menos en el centro, a diferencia de la distribución normal (vea la figura 8.4). Puesto que el valor de  $\sigma$  es desconocido, y se emplea  $S$  para estimarlo, los valores  $t$  son más variables que los valores  $Z$ .

**FIGURA 8.4**

Distribución estandarizada normal y distribución  $t$  para 5 grados de libertad.



Los grados de libertad  $n - 1$  están directamente relacionados con el tamaño de la muestra  $n$ . A medida que el tamaño de la muestra y los grados de libertad se incrementan,  $S$  se vuelve una mejor estimación de  $\sigma$  y la distribución  $t$  gradualmente se acerca a la distribución normal estandarizada hasta que ambas son virtualmente idénticas. Con una muestra de 120 o más,  $S$  estima  $\sigma$  con la suficiente precisión como para que haya poca diferencia entre las distribuciones  $t$  y  $Z$ . Por esta razón, la mayoría de los especialistas en estadística usan  $Z$  en lugar de  $t$  cuando el tamaño de la muestra es mayor a 120.

Como establecimos antes, la distribución  $t$  supone que la variable aleatoria  $X$  se distribuye normalmente. En la práctica, sin embargo, mientras el tamaño de la muestra sea lo suficientemente grande y la población no sea muy sesgada, la distribución  $t$  servirá para estimar la media poblacional cuando  $\sigma$  sea desconocida. Cuando se trabaje con una muestra de tamaño muy pequeño y una distribución poblacional sesgada, la validez del intervalo de confianza es una preocupación. Para evaluar la suposición de normalidad, se evalúa la forma de los datos muestra usando un histograma, una gráfica de tallo y hoja, un diagrama de caja y bigote o una gráfica de probabilidad normal.

Usted encontrará los valores críticos de  $t$  para los grados de libertad adecuados en la tabla para la distribución  $t$  (vea la tabla E.3). Las columnas de la tabla representan el área de la cola superior de la distribución  $t$ . Cada fila representa el valor  $t$  determinado para cada grado de libertad específico. Por ejemplo, con 99 grados de libertad, si se quiere un nivel de confianza del 95%, se encuentra el valor  $t$  apropiado como se muestra en la tabla 8.1. El nivel de confianza del 95% significa que el 2.5% de los valores (un área de .025) se encuentran en cada extremo de la distribución. Buscando en la columna para un área de la cola superior y en la fila correspondiente a 99 grados de libertad, se obtiene un valor crítico para  $t$  de 1.9842. Puesto que  $t$  es una distribución simétrica con una media de 0, si el valor de la cola superior es +1.9842, el valor para el área de la cola inferior (.025 inferior) sería -1.9842. Un valor  $t$  de -1.9842 significa que la probabilidad de que  $t$  sea menor a -1.9842, es de 0.025, o 2.5% (vea la figura 8.5).

$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$

4/2

**TABLA 8.1**

Determinación del valor crítico de la tabla  $t$  para un área de 0.025 en cada cola con 99 grados de libertad.

Grados de libertad	Áreas de la cola superior					
	.25	.10	.05	.025	.01	.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
.	.	.	.	.	.	.
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2902	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259

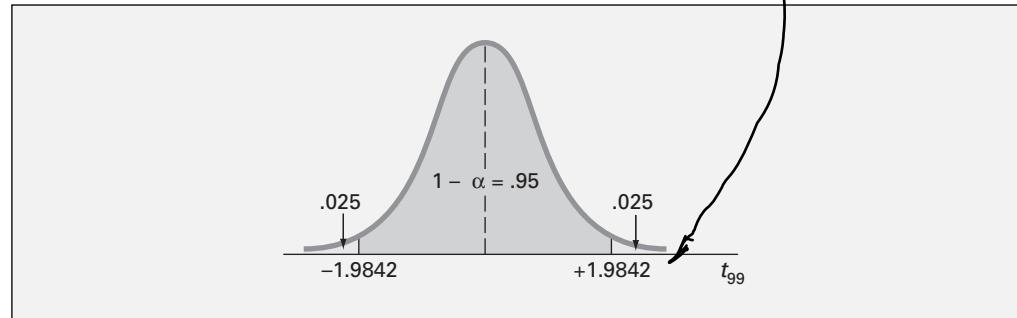
n ~ 1



Fuente: Extraído de la tabla E.3.

**FIGURA 8.5**

Distribución  $t$  para 99 grados de libertad.

**El concepto de grados de libertad**

En el capítulo 3 aprendió que el numerador de la varianza de la muestra  $S^2$  [vea la ecuación (3.9) en la página 82] requiere del cálculo de

$$\sum_{i=1}^n (X_i - \bar{X})^2$$

Para calcular  $S^2$ , primero necesita conocer  $\bar{X}$ . Por lo tanto, sólo  $n - 1$  de los valores de la muestra son libres de variar. Esto significa que tiene  $n - 1$  grados de libertad. Por ejemplo, suponga que una muestra de cinco valores tiene una media de 20. ¿Cuántos valores requiere conocer antes de poder determinar los valores restantes? El hecho de que  $n = 5$  y  $\bar{X} = 20$  también indica que

$$\sum_{i=1}^n X_i = 100$$

ya que

$$\frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

En consecuencia, cuando se conocen cuatro de los valores, el quinto *no* será libre de variar porque la suma debe ser igual a 100. Por ejemplo, si cuatro de los valores son 18, 24, 19 y 16, el quinto valor debe ser 23 para que la suma sea igual a 100.

## El establecimiento del intervalo de confianza

La ecuación (8.2) define la estimación del intervalo de confianza  $(1 - \alpha) \times 100\%$  para la media con  $\sigma$  desconocida.

### INTERVALO DE CONFIANZA PARA LA MEDIA ( $\sigma$ DESCONOCIDA)

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}$$

o

$$\bar{X} - t_{n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} \frac{S}{\sqrt{n}} \quad (8.2)$$

donde  $t_{n-1}$  es el valor crítico de la distribución  $t$  con  $n - 1$  grados de libertad para un área de  $\alpha/2$  en la cola superior.

Para ilustrar la aplicación de la estimación del intervalo de confianza para la media cuando la desviación estándar  $\sigma$  es desconocida, regrese al escenario “Uso de la estadística” de la empresa de remodelaciones Saxon, de la página 238. Seleccione una muestra de 100 facturas de la población de facturas de ventas durante el mes; la media de la muestra de las 100 facturas de ventas es de \$110.27, con una desviación estándar de \$28.95. Para un nivel de confianza del 95%, el valor crítico derivado de la distribución  $t$  (como se muestra en la tabla 8.1) es de 1.9842. Al emplear la ecuación (8.2),

$$\begin{aligned} \bar{X} &\pm t_{n-1} \frac{S}{\sqrt{n}} \\ &= 110.27 \pm (1.9842) \frac{28.95}{\sqrt{100}} \\ &= 110.27 \pm 5.74 \\ &\$104.53 \leq \mu \leq \$116.01 \end{aligned}$$

Se presenta una hoja de trabajo de Excel para estos datos en la figura 8.6.

**FIGURA 8.6**

Hoja de trabajo de Excel para calcular una estimación del intervalo de confianza para la media de la cantidad de facturas de ventas para la empresa de remodelaciones Saxon.

A	B
1 Estimate for the Mean Sales Invoice Amount	
2	
3 Data	
4 Sample Standard Deviation	28.95
5 Sample Mean	110.27
6 Sample Size	100
7 Confidence Level	95%
8	
9 Intermediate Calculations	
10 Standard Error of the Mean	2.8950
11 Degrees of Freedom	99
12 t Value	1.9842
13 Interval Half Width	5.7443
14	
15 Confidence Interval	
16 Interval Lower Limit	104.53
17 Interval Upper Limit	116.01

=B4/RAIZ(B6)  
 =B6 - 1  
 =DISTR.T INV.(1-B7,B11)  
 =B12 \* B10  
 =B5 - B13  
 =B5 + B13

Así, con un nivel de confianza del 95%, usted concluye que la cantidad media de las facturas de ventas está entre \$104.53 y \$116.01. El nivel de confianza del 95% indica que si selecciona todas las muestras posibles de 100 (algo que nunca se hace en la práctica), el 95% de los intervalos desarrollados incluirían la media poblacional en algún lugar dentro del intervalo. La validez de esta estimación del intervalo de confianza depende de la suposición de la normalidad de la distribución de la cantidad de las facturas de ventas. Con una muestra de 100, la suposición de normalidad no es demasiado restrictiva y posiblemente es adecuado el uso de la distribución  $t$ . El ejemplo 8.3 ilustra cómo se construye el intervalo de confianza para la media cuando la desviación estándar de la población es desconocida.

### EJEMPLO 8.3

#### ESTIMACIÓN DE LA FUERZA MEDIA REQUERIDA PARA ROMPER LOS AISLANTES ELÉCTRICOS

Una empresa manufacturera produce aislantes eléctricos. Si los aislantes se rompen al usarse, muy posiblemente tendremos un corto circuito. Para probar la fuerza de los aislantes, se lleva a cabo una prueba destructiva para determinar cuánta fuerza se requiere para romperlos. Se mide la fuerza observando cuántas libras se aplican al aislante antes de que se rompa. La tabla 8.2 lista 30 valores de este experimento. **FORCE** Construya una estimación del intervalo de confianza del 95% para la población media de fuerza requerida para romper el aislante.

**TABLA 8.2**

Fuerza (en libras) requerida para romper el aislante.

1,870	1,728	1,656	1,610	1,634	1,784	1,522	1,696	1,592	1,662
1,866	1,764	1,734	1,662	1,734	1,774	1,550	1,756	1,762	1,866
1,820	1,744	1,788	1,688	1,810	1,752	1,680	1,810	1,652	1,736

**SOLUCIÓN** La figura 8.7 muestra que la media muestral es  $\bar{X} = 1,723.4$  libras y la desviación estándar es  $S = 89.55$  libras. Al utilizar la ecuación (8.2) de la página 246 para construir un intervalo de confianza, se requiere determinar el valor crítico a partir de la tabla  $t$  para un área de 0.025 en cada cola con 29 grados de libertad. En la tabla E.3, se observa que  $t_{29} = 2.0452$ . Así, al considerar que  $\bar{X} = 1,723.4$ ,  $S = 89.55$ ,  $n = 30$  y  $t_{29} = 2.0452$ ,

$$\begin{aligned} \bar{X} &\pm t_{n-1} \frac{S}{\sqrt{n}} \\ &= 1,723.4 \pm (2.0452) \frac{89.55}{\sqrt{30}} \\ &= 1,723.4 \pm 33.44 \\ &1,689.96 \leq \mu \leq 1,756.84 \end{aligned}$$

**FIGURA 8.7**

Estimación en Minitab del intervalo de confianza para la media de cantidad de fuerza requerida para romper los aislantes eléctricos.

#### One-Sample T: Force

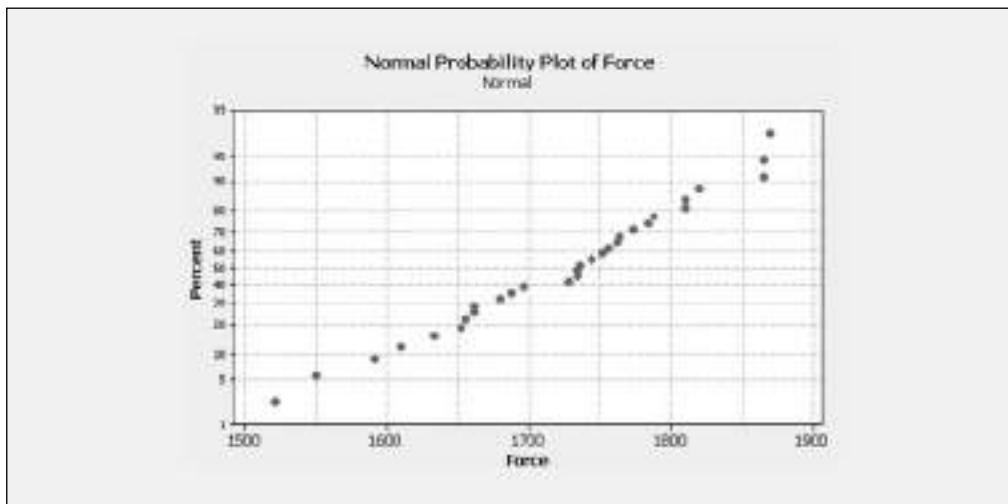
Variable	N	Mean	StDev	SE Mean	95% CI
Force	30	1723.40	89.55	16.35	(1689.96, 1756.84)

Usted concluye con un nivel de confianza del 95% que la fuerza media requerida para la población de aislantes está entre 1,689.96 y 1,756.84 libras. La validez de esta estimación del intervalo de confianza depende de la suposición de que la fuerza requerida se distribuye normalmente. Sin embargo, recuerde que es factible relajar ligeramente esta suposición cuando trabaja con muestras de

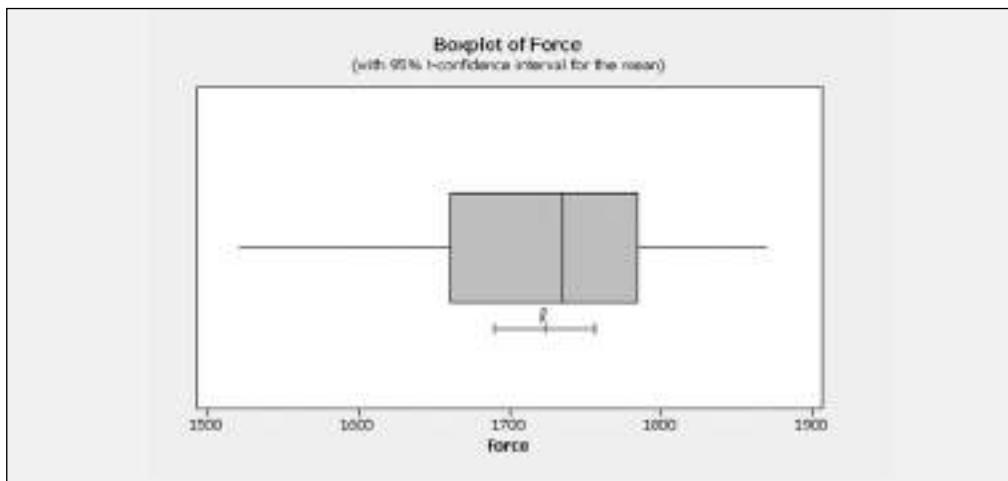
mayor tamaño. Por tanto, con una muestra de 30, puede usar una distribución  $t$  aun cuando la cantidad de fuerza requerida esté ligeramente sesgada. Para la gráfica de probabilidad normal que muestra la figura 8.8, o en la gráfica de caja y bigote de la figura 8.9, la cantidad de fuerza requerida parece ligeramente sesgada. Por tanto, la distribución  $t$  es apropiada para estos datos.

**FIGURA 8.8**

Gráfica de probabilidad normal de Minitab para la cantidad de fuerza requerida para romper los aislantes eléctricos.

**FIGURA 8.9**

Gráfica de caja y bigote de Minitab para la cantidad de fuerza requerida para romper los aislantes eléctricos.



## PROBLEMAS PARA LA SECCIÓN 8.2

### Aprendizaje básico

**ASISTENCIA de PH Grade** **8.10** Determine el valor crítico de  $t$  en cada una de las siguientes circunstancias:

- $1 - \alpha = 0.95, n = 10$ .
- $1 - \alpha = 0.99, n = 10$ .
- $1 - \alpha = 0.95, n = 32$ .
- $1 - \alpha = 0.95, n = 65$ .
- $1 - \alpha = 0.90, n = 16$ .

**ASISTENCIA de PH Grade** **8.11** Si  $\bar{X} = 75, S = 24, n = 36$ , y suponiendo que la población se distribuye normalmente, construya una estimación del intervalo de confianza del 95% de la media poblacional  $\mu$ .

**ASISTENCIA de PH Grade**

**8.12** Si  $\bar{X} = 50, S = 15, n = 16$ , y suponiendo que la población se distribuye normalmente, construya una estimación del intervalo de confianza del 99% de la media poblacional  $\mu$ .

**8.13** Construya una estimación del intervalo de confianza del 95% para la media poblacional, basada en cada uno de los siguientes conjuntos de datos, suponiendo que la población se distribuye normalmente:

Conjunto 1: 1, 1, 1, 1, 8, 8, 8, 8

Conjunto 2: 1, 2, 3, 4, 5, 6, 7, 8

Explique por qué estos conjuntos de datos tienen diferentes intervalos de confianza aun cuando tienen la misma media y el mismo rango.

**8.14** Construya un intervalo de confianza del 95% para la media poblacional, basado en los números 1, 2, 3, 4, 5, 6 y 20. Cambie el número 20 por el 7 y recalcule el intervalo de confianza. Usando estos resultados, describa el efecto de un valor atípico (es decir, un valor extremo) en un intervalo de confianza.

### Aplicación de conceptos

**Puede resolver los problemas 8.15 a 8.22 con o sin Excel, Minitab o SPSS. Debe usar Excel, Minitab o SPSS para resolver el problema 8.23.**

ASISTENCIA  
de PH Grade

AUTO  
Examen

**8.15** El propietario de una papelería desea estimar la media del valor al menudeo de las tarjetas de felicitación que la tienda tiene en su inventario. Una muestra aleatoria de 20 tarjetas de felicitación indica una media de valor de \$1.67 y una desviación estándar de \$0.32.

- Suponiendo una distribución normal, construya una estimación del intervalo de confianza del 95% para la media del valor de todas las tarjetas de felicitación en el inventario de la tienda.
- ¿Cómo podrían ser útiles los resultados del inciso a) para ayudar al dueño de la tienda a estimar el valor total de su inventario?

**8.16** El hospital Southside en Bay Shore, de Nueva York, generalmente lleva a cabo pruebas de estrés para estudiar el músculo cardíaco después de que una persona ha sufrido un ataque al corazón. Los miembros del departamento de imagen diagnóstica dirigen un proyecto de mejoramiento para tratar de reducir el tiempo de procesamiento de las pruebas de estrés. El tiempo de procesamiento se define como el tiempo que pasa desde que se ordena la prueba hasta que el radiólogo firma los resultados. Inicialmente la media del tiempo de procesamiento para los resultados de una prueba de estrés era de 68 horas. Después de incorporar cambios al proceso de pruebas de esfuerzo, el equipo de mejoramiento de calidad recolectó una muestra de 50 tiempos de procesamiento. En esta muestra, la media de procesamiento fue de 32 horas con una desviación estándar de 9 horas (Eric Godin, Dennis Raven, Carolyn Sweetapple y Frank R. Del Guidice, "Faster Test Results", *Quality Progress*, enero de 2004, 37(1):33-39).

- Construya un intervalo de confianza del 95% para la media poblacional del tiempo de procesamiento.
- Interprete el intervalo construido en el inciso a).
- ¿Cree que el proyecto de mejoramiento de calidad fue exitoso? Explique por qué.

ASISTENCIA  
de PH Grade

**8.17** El Departamento del Transporte de Estados Unidos requiere a los fabricantes de neumáticos proporcionar información sobre el desempeño de los neumáticos en la pared interna para informar mejor a los futuros clientes mientras toman su decisión de compra. Una medida muy importante en el desempeño de los neumáticos es el índice de la banda de rodamiento, el cual indica la resistencia del neumático al rodamiento en comparación con un neumático graduado con una base de 100. Esto significa que un neumático con un grado de 200 debería durar al menos el doble de tiempo, en promedio, que un neumático graduado con una base de 100. Una organización de consumidores desea estimar el índice real de rodamiento para una marca de neumáticos graduada 200 que produce un determinado fabricante. Una muestra aleatoria de

$n = 18$  reporta una media muestral de índice de banda de rodamiento de 195.3 y una desviación estándar de 21.4.

- Suponiendo que la población de índices de banda de rodamiento se distribuye normalmente, construya una estimación del intervalo de confianza de la media poblacional del índice de rodamiento para neumáticos hechos por este fabricante bajo esta marca.
- ¿Cree que la organización de consumidores debería acusar al fabricante de producir neumáticos que no cumplen con la información de desempeño proporcionada en la pared interna del neumático? Explique su respuesta.
- Explique por qué un índice de rodamiento observado de 210 para un neumático específico es usual, aun cuando está fuera del intervalo de confianza desarrollado en el inciso a).

**8.18** Los siguientes datos representan la tarifa en dólares para los cheques sin fondos para una muestra de 23 bancos con clientes que depositan directamente y que mantienen un saldo de \$100. **BANKCOST1**

26 28 20 20 21 22 25 25 18 25 15 20  
18 20 25 25 22 30 30 30 15 20 29

Fuente: "The New Face of Banking", Copyright © 2000 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con el permiso de Consumer Reports, junio de 2000.

- Construya un intervalo de confianza del 95% para la media poblacional de cheques sin fondos.
- Interprete el intervalo construido en el inciso a).

**8.19** Los siguientes datos representan la tarifa del servicio mensual en dólares, si el saldo en la cuenta de un cliente cae debajo del saldo mínimo requerido de \$1,500 para una muestra de 26 bancos con clientes con cuentas de depósito directo. **BANKCOST2**

12 8 5 5 6 6 10 10 9 7 10 7 7  
5 0 10 6 9 12 0 5 10 8 5 5 9

Fuente: "The New Face of Banking", Copyright © 2000 por Consumers Union of U.S., Inc., Yonkers, NY 10703-1057. Adaptado con el permiso de Consumer Reports, junio de 2000.

- Construya un intervalo de confianza del 95% para la media poblacional de la tarifa en dólares del servicio si el saldo de un cliente baja del saldo mínimo requerido.
- Interprete el intervalo construido en el inciso a).

**8.20** Una de las principales medidas en el servicio de calidad proporcionado por cualquier organización es la velocidad con la que se responde a las quejas del cliente. Una gran empresa familiar de muebles y pisos, incluyendo alfombras, ha experimentado una gran expansión durante los últimos años. El departamento de pisos, en particular, ha crecido de dos cuadrillas de instalación a un supervisor de instalación, un medidor y 15 cuadrillas de instalación. El año pasado hubo 50 quejas relacionadas con la instalación de alfombras. Los siguientes datos **FURNITURE** representan el número de días que pasaron desde que se recibió la queja hasta que ésta se solucionó.

54 5 35 137 31 27 152 2 123 81 74 27  
11 19 126 110 110 29 61 35 94 31 26 5  
12 4 165 32 29 28 29 26 25 1 14 13  
13 10 5 27 4 52 30 22 36 26 20 23  
33 68

- a. Construya una estimación del intervalo de confianza del 95% de la media del número de días entre que se recibe la queja y se resuelve.
- b. ¿Qué suposición debe hacer sobre la distribución poblacional en el inciso a)?
- c. ¿Cree que la suposición del inciso b) es infringida seriamente? Explique su respuesta.
- d. ¿Qué efecto podría tener su conclusión en el inciso c) sobre la validez de los resultados en el inciso a)?

**8.21** En el estado de Nueva York se permite a la banca de ahorro vender un seguro de vida llamado Savings Bank Life Insurance (SBLI). El proceso de aprobación consiste en suscribir, lo que incluye revisar la solicitud, realizar un control del centro de información médica, realizar posibles peticiones para información médica adicional y exámenes médicos, y una fase de compilación de póliza, donde las páginas de la póliza se generan y se envían al banco para su entrega. Para el banco, la capacidad de entrega oportuna de las pólizas aprobadas a los clientes es esencial para la rentabilidad de este servicio. Se seleccionó una muestra aleatoria de 27 pólizas aprobadas durante el periodo de 1 mes INSURANCE y el tiempo de procesamiento total en días registrado:

73 19 16 64 28 28 31 90 60 56 31 56 22 18  
45 48 17 17 17 91 92 63 50 51 69 16 17

- a. Construya una estimación del intervalo de confianza del 95% de la media del tiempo de procesamiento.
- b. ¿Qué suposición debe hacerse acerca de la distribución de la población en el inciso a)?
- c. ¿Cree que la suposición hecha en el inciso b) es infringida seriamente? Explique su respuesta.

**8.22** Los siguientes datos representan el costo diario de hotel y de renta de un automóvil en 20 ciudades de Estados Unidos durante una semana en el mes de octubre de 2003. HOTEL-CAR

Ciudad	Hotel	Automóviles
San Francisco	205	47
Los Angeles	179	41
Seattle	185	49
Phoenix	210	38
Denver	128	32
Dallas	145	48
Houston	177	49
Minneapolis	117	41
Chicago	221	56

Ciudad	Hotel	Automóviles
St. Louis	159	41
New Orleans	205	50
Detroit	128	32
Cleveland	165	34
Atlanta	180	46
Orlando	198	41
Miami	158	40
Pittsburgh	132	39
Boston	283	67
New York	269	69
Washington D.C.	204	40

Fuente: *The Wall Street Journal*, 10 de octubre, 2003. W4.

- a. Construya un intervalo de confianza del 95% para la media poblacional del precio del hotel.
- b. Construya un intervalo de confianza del 95% para la media poblacional del precio de la renta de automóvil.
- c. ¿Qué suposición requiere hacer acerca de las poblaciones de interés para construir los intervalos en los incisos a) y b)?
- d. A partir de los datos presentados, ¿cree que la suposición necesaria en los incisos a) y b) es válida? Explique por qué.

**8.23** Una de las operaciones de un molino consiste en cortar piezas de acero en partes que posteriormente se usarán como marco para los asientos delanteros de los automóviles en una planta. El acero se corta con una sierra de diamante y se requiere que las partes resultantes sean más o menos de 0.005 pulgadas de la longitud especificada por la empresa automotriz. Los datos provienen de una muestra de 100 partes de acero STEEL. La medida que se reporta es la diferencia en pulgadas entre la longitud real de la parte de acero, medida con láser, y la longitud especificada de la parte de acero. Por ejemplo, la primera observación, -0.002, representa una parte de acero 0.002 pulgadas más corta de la longitud especificada.

- a. Construya una estimación del intervalo de confianza del 95% de la media de la diferencia entre la longitud real de la parte de acero y la longitud especificada de la parte de acero.
- b. ¿Qué suposición debe hacer acerca de la distribución poblacional en el inciso a)?
- c. ¿Cree que la suposición hecha en el inciso b) es infringida seriamente? Explique su respuesta.
- d. Compare las conclusiones del inciso a) con las del problema 2.23 en la página 42.

### 8.3 ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN

Esta sección extiende el concepto de intervalo de confianza a los datos categóricos. Aquí nos preocuaremos de la estimación de la proporción de elementos en una población que tiene ciertas características de interés. La proporción desconocida de la población, se representa con la letra griega  $\pi$ . La estimación puntual para  $\pi$  es la proporción de la muestra,  $p = X/n$ , donde  $n$  es el tamaño de la muestra y  $X$  es el número de elementos en la muestra que tienen la característica de interés. La ecuación (8.3) define la estimación del intervalo de confianza para la proporción de la población.

## ESTIMACIÓN DEL INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN

$$p \pm Z \sqrt{\frac{p(1-p)}{n}}$$

o

$$p - Z \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + Z \sqrt{\frac{p(1-p)}{n}} \quad (8.3)$$

donde  $p$  = proporción de la muestra =  $\frac{X}{n}$  =  $\frac{\text{número de elementos con característica}}{\text{tamaño de la muestra}}$

$\pi$  = proporción de la población

$Z$  = valor crítico para la distribución normal estandarizada

$n$  = tamaño de la muestra

suponiendo que tanto  $X$  como  $n - X$  son mayores que 5.

Se puede usar la estimación del intervalo de confianza de la proporción definida en la ecuación (8.3) para estimar la proporción de las facturas de ventas que contienen errores (vea el escenario “Uso de la estadística” en la página 238). Suponga que en una muestra de 100 facturas de ventas, 10 contienen errores. Así entonces, para estos datos,  $p = X/n = 10/100 = 0.10$ . Al emplear la ecuación (8.3) y  $Z = 1.96$  para un 95% de confianza,

$$\begin{aligned} p \pm Z \sqrt{\frac{p(1-p)}{n}} \\ = 0.10 \pm (1.96) \sqrt{\frac{(0.10)(0.90)}{100}} \\ = 0.10 \pm (1.96)(0.03) \\ = 0.10 \pm 0.0588 \\ 0.0412 \leq \pi \leq 0.1588 \end{aligned}$$

Así, usted tiene un nivel de confianza del 95% de que entre el 4.12 y el 15.88% de todas las facturas de ventas contienen errores. La figura 8.10 muestra una hoja de trabajo de Excel para estos datos, mientras que la figura 8.11 ilustra un resultado de Minitab.

**FIGURA 8.10**

Hoja de trabajo de Excel para formar una estimación del intervalo de confianza para la proporción de facturas de ventas que contienen errores.

A	B
1	<b>Proportion of In-Error Sales Invoices</b>
2	
3	<b>Data</b>
4	<b>Sample Size</b> 100
5	<b>Number of Successes</b> 10
6	<b>Confidence Level</b> 95%
7	
8	<b>Intermediate Calculations</b>
9	<b>Sample Proportion</b> 0.1
10	<b>Z Value</b> 1.9600
11	<b>Standard Error of the Proportion</b> 0.03
12	<b>Interval Half-Width</b> 0.0588
13	
14	<b>Confidence Interval</b>
15	<b>Interval Lower Limit</b> 0.0412
16	<b>Interval Upper Limit</b> 0.1588

$=B5/B4$   
 $=DISTR.NORM.ESTAND.INV((1 - B6)/2)$   
 $=RAIZ(B9 * (1 - B6)/B4)$   
 $=ABS(B10 * B11)$   
  
 $=B9 - B12$   
 $=B9 - B12$

Sample	X	N	Sample p	95% CI
1	10	100	0.100000	(0.041201, 0.158799)

**FIGURA 8.11** Estimación en Minitab del intervalo de confianza para la proporción de facturas de ventas que contienen errores.

El ejemplo 8.4 ilustra otra aplicación de la estimación del intervalo de confianza para la proporción.

### EJEMPLO 8.4

### ESTIMACIÓN DE LA PROPORCIÓN DE PERIÓDICOS DEFECTUOSOS IMPRESOS

El editor de un periódico desea estimar la proporción de periódicos impresos con algún defecto, tal como borraduras en exceso, disposición errónea de las hojas, páginas faltantes o duplicadas. Se selecciona una muestra aleatoria de 200 periódicos, 35 de ellos contienen algún tipo de defecto. Realice e interprete un intervalo de confianza del 90% para la proporción de periódicos impresos durante el día que tienen defectos.

**SOLUCIÓN** Se emplea la ecuación (8.3):

$$\begin{aligned}
 p &= \frac{35}{200} = 0.175, \text{ y con un nivel de confianza del } 90\% Z = 1.645 \\
 p \pm Z \sqrt{\frac{p(1-p)}{n}} &= 0.175 \pm (1.645) \sqrt{\frac{(0.175)(0.825)}{200}} \\
 &= 0.175 \pm (1.645)(0.0269) \\
 &= 0.175 \pm 0.0442 \\
 0.1308 \leq \pi &\leq 0.2192
 \end{aligned}$$

Usted concluye con un 90% de confianza que entre el 13.08 y el 21.92% de los periódicos impresos en el día tienen algún defecto.

La ecuación (8.3) contiene un estadístico  $Z$  ya que se puede usar la distribución normal para aproximar la distribución binomial cuando el tamaño de la muestra es lo suficientemente grande. En el ejemplo 8.4, el intervalo de confianza usando  $Z$  proporciona una excelente aproximación para la proporción de la población porque tanto  $X$  como  $n - X$  son mayores que 5. Sin embargo, si no se tiene un tamaño de muestra lo suficientemente grande, entonces se debería usar la distribución binomial en lugar de la ecuación (8.3) (vea las referencias 1, 2 y 7). Fisher y Yates (vea la referencia 2) han tabulado intervalos de confianza exactos para diferentes tamaños de muestras y para proporciones de éxitos, y pueden calcularse con Minitab.

## PROBLEMAS PARA LA SECCIÓN 8.3

### Aprendizaje básico

**ASISTENCIA de PH Grade** **8.24** Si  $n = 200$  y  $X = 50$ , construya una estimación del intervalo de confianza del 95% para la proporción de la población.

**ASISTENCIA de PH Grade** **8.25** Si  $n = 400$  y  $X = 25$ , construya una estimación del intervalo de confianza del 99% para la proporción de la población.

## Aplicación de conceptos

ASISTENCIA  
de PH Grade

AUTO  
Examen

**8.26** Una empresa telefónica desea estimar la proporción de hogares en los que se contrataría una línea telefónica adicional. Se seleccionó una muestra aleatoria de 500 hogares. Los resultados indican que a un costo reducido, 135 de los hogares contratarían una línea telefónica adicional.

- Construya una estimación del intervalo de confianza del 99% de la proporción poblacional de hogares que contratarían una línea telefónica adicional.
- ¿Cómo podría el gerente a cargo de los programas promocionales relacionados con los clientes residenciales, usar los resultados del inciso a)?

**8.27** De acuerdo con el Center for Work-Life Policy, una encuesta realizada con 500 mujeres con altos niveles educativos que abandonaron sus carreras por razones familiares, indicó que el 66% de ellas deseaban regresar al trabajo (Anne Marie Chaker y Hillary Stout, "After Years Off, Women Struggle to Revive Careers", *The Wall Street Journal*, mayo 6, 2004, A1).

- Construya un intervalo de confianza del 95% para la proporción poblacional de mujeres con altos niveles educativos que abandonaron sus carreras por razones familiares y que desean regresar al trabajo.
- Interprete el intervalo en el inciso a).

**8.28** Millones de estadounidenses organizan sus planes de viaje en Internet ("Travelers Head Online", *USA Today Snapshots*, 22 de julio, 2003). En una encuesta reciente, se reportó que el 77% compra boletos de avión en Internet. Suponga que la encuesta se basó en 1,000 sujetos que respondieron.

- Construya una estimación del intervalo de confianza del 95% para la proporción poblacional de estadounidenses que compran boletos de avión en Internet.
- Construya una estimación del intervalo de confianza del 90% para la proporción poblacional de estadounidenses que compran boletos de avión en Internet.
- ¿Cuál intervalo es más amplio? Explique por qué esto es cierto.

**8.29** El número de consumidores de edad avanzada en Estados Unidos se incrementa, de manera que se están convirtiendo en una fuerza económica mayor. Muchos se sienten abrumados cuando se enfrentan a la tarea de seleccionar inversiones, servicios bancarios, proveedores de cuidado de la salud o de servicios telefónicos. Se realizó una encuesta telefónica de 1,900 consumidores de edad avanzada; el 27% de ellos afirmaron que no tenían el suficiente tiempo para ser buenos administradores de dinero ("Seniors Confused by Financial Choices-Study", [msnbc.com](http://msnbc.com), mayo 6, 2004).

- Construya un intervalo de confianza del 95% para la proporción poblacional de consumidores de edad avanzada que creen no tener tiempo suficiente para ser buenos administradores de dinero.
- Interprete el intervalo en el inciso a).

**8.30** Para estudiar el problema del uso de celulares mientras se maneja ("Drivers Using Cell Phones Have Problems", *USA Today*, 16 de mayo, 2001, 1A), se llevó a cabo una encuesta entre conductores que utilizan celulares mientras manejan. En la encuesta, el 46% de quienes

respondieron reportaron haber tenido que virar bruscamente y el 10% conocía a alguien que había chocado mientras hablaba por celular. Suponga que la encuesta se basó en 500 personas que respondieron.

- Construya un intervalo de confianza del 95% para la proporción de todos los conductores que han virado bruscamente.
- Construya un intervalo de confianza del 95% para la proporción de todos los conductores que conocían a alguien que chocó mientras hablaba por celular.

**8.31** Conforme se incrementan los costos de seguros de gastos médicos y aumenta el número de empleados con discapacidad, más compañías despiden a estos empleados. Una encuesta que se realizó entre 723 empleados encontró que 195 empleados fueron despedidos en cuanto se les diagnosticó una discapacidad crónica (J. Pereira, "To Save on Health-care Costs, Firms Fire Disabled Workers", *The Wall Street Journal*, 14 de julio, 2003, A1-A7).

- Construya un intervalo de confianza del 95% para la proporción de empleadores que despidieron empleados en cuanto éstos manifestaron una discapacidad crónica.
- Construya un intervalo de confianza del 99% para la proporción de empleadores que despidieron empleados en cuanto éstos manifestaron una discapacidad crónica.
- ¿Cuál intervalo es el más amplio? Explique por qué esto es verdad.

**8.32** La unidad Clinique de los cosméticos Estée Lauder realizó una encuesta entre mujeres trabajadoras en Norteamérica. De 1,000 mujeres encuestadas, el 55% pensaba que las empresas deberían reservar los puestos durante seis meses o menos para aquellas con permiso de maternidad, y el 45% consideraba que deberían reservar sus puestos durante más de seis meses ("Work Week", *The Wall Street Journal*, 11 de septiembre, 2001, A1).

- Construya un intervalo de confianza del 95% para la proporción de las mujeres trabajadoras en Norteamérica quienes creen que las empresas deberían reservar los puestos durante seis meses o menos para aquellas con permiso de maternidad.
- Interprete el intervalo construido en a).

**8.33** Un gran número de empresas tratan de reducir el costo de los medicamentos prescritos, solicitando a sus empleados que los compren a través de un programa obligatorio de orden por correo. En una encuesta realizada entre 600 empleados, 126 indicaron que tienen un programa obligatorio de orden por correo, o bien, adoptarían uno para fines de 2004 (Barbara Martínez, "Forcing Employees to Buy Drugs Via Mail", *The Wall Street Journal*, 18 de febrero, 2004, D1).

- Construya un intervalo de confianza del 95% para la proporción de la población de empleados que tienen un programa obligatorio de orden por correo o que adoptarían uno para finales de 2004.
- Construya un intervalo de confianza del 99% para la proporción de la población de empleados que tienen un programa obligatorio de orden por correo o que adoptarían uno para finales de 2004.
- Interprete los intervalos en los incisos a) y b).
- Discuta el efecto en la estimación del intervalo de confianza al modificar el nivel de confianza.

## 8.4 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA

En cada ejemplo de la estimación del intervalo de confianza, usted seleccionó un tamaño de muestra sin considerar la amplitud del intervalo de confianza. En el mundo de los negocios, determinar el tamaño adecuado de la muestra es un procedimiento complicado, ya que está sujeto a las restricciones de presupuesto, tiempo y cantidad aceptada de error de muestreo. Si en el ejemplo de la empresa de remodelaciones Saxon usted desea estimar la media de la cantidad de ventas en dólares de las facturas de ventas o la proporción de facturas de ventas que contienen errores, debe determinar por anticipado cuán grande sería el error de muestreo para estimar cada uno de los parámetros. También debe determinar por anticipado el nivel del intervalo de confianza a usar al estimar el parámetro poblacional.

### Determinación del tamaño de la muestra para la media

Para desarrollar una fórmula que permita determinar el tamaño apropiado de la muestra para construir una estimación del intervalo de confianza para la media, recuerde la ecuación (8.1) en la página 241:

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

La cantidad sumada o sustraída de  $\bar{X}$  es igual a la mitad de la amplitud del intervalo. Esta cantidad representa la cantidad de imprecisión en la estimación que resulta del error de muestreo. El **error de muestreo**<sup>1</sup>  $e$  se define como

$$e = Z \frac{\sigma}{\sqrt{n}}$$

Al resolver  $n$  se obtiene el tamaño de muestra necesario para construir una estimación del intervalo de confianza adecuado para la media. “Adecuado” significa que el intervalo resultante tendrá una cantidad aceptable de error de muestreo.

#### DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA LA MEDIA

El tamaño de la muestra  $n$  es igual al producto del valor  $Z$  al cuadrado y de la varianza  $\sigma^2$ , dividido por el error de muestreo  $e$  cuadrada.

$$n = \frac{Z^2 \sigma^2}{e^2} \quad (8.4)$$

Para determinar el tamaño de la muestra debe conocer tres factores:

1. El nivel de confianza deseado, el cual determina el valor de  $Z$ , el valor crítico de la distribución normal estandarizada.<sup>2</sup>
2. El error de muestreo aceptable  $e$ .
3. La desviación estándar  $\sigma$ .

En algunas relaciones de negocio a negocio que requieren de la estimación de parámetros importantes, los contratos legales son los que especifican los niveles de error aceptables y el nivel de confianza requerido. Para las empresas del sector alimentario o medicinal, las regulaciones gubernamentales a menudo especifican los errores de muestreo y los niveles de confianza. En general, sin embargo, no es tan fácil especificar los dos factores necesarios para determinar el tamaño de la muestra. ¿Cómo se determina el nivel de confianza y el error de muestreo? Por lo general, sólo por un experto en la materia (es decir, el individuo más familiarizado con las variables estudiadas) es quien responde a estas preguntas. Aunque el 95% es el nivel de confianza más usado, si desea mayor confianza, entonces el 99% podría ser más adecuado; si considera aceptable un menor nivel de confianza, entonces podría usar el 90%. Para el error de muestreo no debería pensar qué tanto error de muestreo le

<sup>1</sup>En este contexto, algunos especialistas en estadística se refieren a  $e$  como “margen de error”.

<sup>2</sup>Se usa  $Z$  en lugar de  $t$  porque, para determinar el valor crítico de  $t$  se requiere conocer el tamaño de la muestra, pero aún no se conoce. Para la mayoría de los estudios, el tamaño de la muestra necesario es lo suficientemente grande como para que la distribución normal estandarizada sea una buena aproximación de la distribución  $t$ .

gustaría tener (usted no desea realmente tener errores), sino qué tanto podría tolerar al sacar las conclusiones de estos datos.

Además de especificar el nivel de confianza y el error de muestreo, usted requiere una estimación de la desviación estándar. Por desgracia, rara vez conoce la desviación estándar poblacional  $\sigma$ . En algunas ocasiones es posible estimar la desviación estándar a partir de datos pasados. En otras situaciones, puede hacer una cuidadosa conjectura tomando en cuenta el rango y la distribución de la variable. Por ejemplo, si supone que hay una distribución normal, el rango es aproximadamente igual a  $6\sigma$  (es decir,  $\pm 3\sigma$  alrededor de la media), por lo que puede estimar  $\sigma$  como el rango dividido por 6. Si no es posible estimar la  $\sigma$  de esta forma, tiene la opción de llevar a cabo un estudio a pequeña escala y estimar la desviación estándar a partir de los datos resultantes.

Para investigar cómo determinar el tamaño de la muestra necesario para la estimación de la media poblacional, vuelva a considerar la auditoría de la empresa de remodelaciones Saxon. En la sección 8.2, usted seleccionó una muestra de 100 facturas de ventas y desarrolló una estimación del intervalo de confianza del 95% de la media poblacional de la cantidad registrada en las facturas de ventas. ¿Cómo determinó este tamaño de muestra? ¿Debería haber seleccionado un tamaño de muestra diferente?

Suponga que, después de consultar con los funcionarios de la empresa, determina que desea un error de muestreo no mayor a  $\pm \$5$ , junto con un nivel de confianza del 95%. Los datos pasados indican que la desviación estándar de la cantidad de ventas es aproximadamente de \$25. Entonces,  $e = \$5$ ,  $\sigma = \$25$  y  $Z = 1.96$  (para un nivel de confianza del 95%). Al utilizar la ecuación (8.4),

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.96)^2 (25)^2}{(5)^2} = 96.04$$

Como la regla general es satisfacer ligeramente de más el criterio redondeando el tamaño de la muestra hacia arriba hasta el siguiente número entero, usted debería seleccionar un tamaño de muestra de 97. Así pues, el tamaño de la muestra 100 considerado en la página 246 se acerca a lo requerido para satisfacer las necesidades de la empresa con base en la estimación de la desviación estándar, el nivel de confianza deseado y el error de muestreo. Puesto que la desviación estándar calculada es ligeramente mayor a la esperada, \$28.95 comparada con \$25.00, el intervalo de confianza es ligeramente más amplio que el deseado. La figura 8.12 ilustra una hoja de trabajo de Excel usada para determinar el tamaño de la muestra.

**FIGURA 8.12**

Hoja de trabajo de Excel para determinar el tamaño de la muestra para la estimación de la media de la cantidad de ventas facturadas para la empresa de remodelaciones Saxon.

	A	B
1	For the Mean Sales Invoice Amount	
2		
3	Data	
4	Population Standard Deviation	25
5	Sampling Error	5
6	Confidence Level	95%
7		
8	Intermediate Calculations	
9	Z Value	-1.9600
10	Calculated Sample Size	96.0365
11		
12	Result	
13	Sample Size Needed	97

=NORMSINV((1 - B6)/2)  
 =(B9 \* B4)/B5)^2  
 =ROUNDUP(B10, 0)

El ejemplo 8.5 ilustra otra aplicación de cómo determinar el tamaño de la muestra necesario para desarrollar un intervalo de confianza para la media.

### EJEMPLO 8.5

### DETERMINAR EL TAMAÑO DE LA MUESTRA PARA LA MEDIA

Regresando al ejemplo 8.3 de la página 247, suponga que desea estimar la media poblacional de la fuerza requerida para romper el aislante para que caiga dentro de  $\pm 25$  libras con un 95% de confianza. Con base en un estudio realizado el año anterior, usted cree que la desviación estándar es de 100 libras. Encuentre el tamaño requerido de la muestra.

**SOLUCIÓN** Mediante la ecuación (8.4) en la página 254 y con  $e = 25$ ,  $\sigma = 100$  y  $Z = 1.96$  para un nivel de confianza del 95%

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.96)^2 (100)^2}{(25)^2} = 61.47$$

Por lo tanto, el tamaño de la muestra que se debe seleccionar es de 62 aislantes, porque la regla general para determinar el tamaño de la muestra es siempre redondear hacia arriba al siguiente valor entero para satisfacer ligeramente de más el criterio deseado.

Obtendremos un error de muestreo real ligeramente mayor que 25 si la desviación estándar de la muestra calculada en esta muestra de 62 es mayor a 100, y ligeramente menor si la desviación estándar de la muestra es menor a 100.

### Determinación del tamaño de la muestra para la proporción

Hasta ahora en esta sección usted ha aprendido cómo determinar el tamaño de muestra necesario para estimar la media poblacional. Ahora, suponga que desea determinar el tamaño de muestra necesario para estimar la proporción de facturas de ventas de la empresa de remodelaciones Saxon que contienen errores.

Para determinar el tamaño de la muestra necesario para estimar la proporción poblacional ( $\pi$ ), utilice un método similar al método para calcular la media poblacional. Recuerde que al desarrollar el tamaño de la muestra para un intervalo de confianza para la media, el error de muestreo se define por:

$$e = Z \frac{\sigma}{\sqrt{n}}$$

Cuando estima la proporción, reemplace  $\sigma$  por  $\sqrt{\pi(1 - \pi)}$ . Así, el error de muestreo es:

$$e = Z \sqrt{\frac{\pi(1 - \pi)}{n}}$$

Al resolver  $n$  usted obtiene el tamaño de la muestra necesario para desarrollar la estimación del intervalo de confianza para la proporción.

#### DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA UNA PROPORCIÓN

El tamaño de la muestra  $n$  es igual al valor  $Z$  al cuadrado multiplicado por la proporción poblacional  $\pi$ , multiplicado por 1 menos la proporción poblacional  $\pi$ , dividido por el error de muestreo  $e$  al cuadrado.

$$n = \frac{Z^2 \pi(1 - \pi)}{e^2} \quad (8.5)$$

Para determinar el tamaño de la muestra debe conocer tres factores:

1. El nivel de confianza deseado, que determina el valor de  $Z$ , el valor crítico para la distribución normal estandarizada.
2. El error de muestreo  $e$  aceptable.
3. La proporción poblacional  $\pi$ .

En la práctica, seleccionar estas cantidades requiere de cierta planeación. Una vez que determine el nivel de confianza deseado, podrá encontrar el valor  $Z$  adecuado para la distribución normal estandarizada. El error de muestreo  $e$  indica la cantidad de error que está dispuesto a tolerar al estimar la

proporción poblacional. La tercera cantidad  $\pi$  es en realidad el parámetro poblacional que desea estimar! Entonces, ¿cómo se establece un valor para el mismo objeto del que estamos tomando una muestra para poder determinarlo?

Tenemos aquí dos alternativas. En muchas situaciones se cuenta con información anterior o experiencias relevantes que proporcionan una estimación cuidadosa de  $\pi$ . O, si no se cuenta con información anterior o experiencias relevantes, se trata de proporcionar un valor para  $\pi$  que nunca *subestime* el tamaño de muestra necesario. Con referencia a la ecuación (8.5), observe que la cantidad  $\pi(1 - \pi)$  aparece en el numerador. Por eso, se requiere determinar el valor de  $\pi$  que haga la cantidad  $\pi(1 - \pi)$  lo más grande posible. Cuando  $\pi = 0.5$ , el producto  $\pi(1 - \pi)$  logra su resultado máximo. Para mostrar esto, los valores de  $\pi$  junto con los productos que los acompañan de  $\pi(1 - \pi)$  son como sigue:

Cuando  $\pi = 0.9$ , entonces  $\pi(1 - \pi) = (0.9)(0.1) = 0.09$

Cuando  $\pi = 0.7$ , entonces  $\pi(1 - \pi) = (0.7)(0.3) = 0.21$

Cuando  $\pi = 0.5$ , entonces  $\pi(1 - \pi) = (0.5)(0.5) = 0.25$

Cuando  $\pi = 0.3$ , entonces  $\pi(1 - \pi) = (0.3)(0.7) = 0.21$

Cuando  $\pi = 0.1$ , entonces  $\pi(1 - \pi) = (0.1)(0.9) = 0.09$

Por lo tanto, cuando no se tiene conocimiento previo o una estimación de la proporción poblacional  $\pi$ , se debería usar  $\pi = 0.5$  para determinar el tamaño de la muestra. Esto produce el tamaño de muestra más grande posible y deriva en el mayor costo posible del muestreo. Al utilizar  $\pi = 0.5$  podría sobreestimarse el tamaño de muestra necesario porque se usa la proporción de muestra real al desarrollar el intervalo de confianza. Obtendrá un intervalo de confianza más estrecho que el que pretendía originalmente, si la muestra de la proporción real es diferente a 0.5. Una precisión más alta deriva en tener que invertir más tiempo y dinero para conseguir una muestra de mayor tamaño.

Volviendo al escenario de “Uso de la estadística” de la empresa de remodelaciones Saxon, suponga que los procedimientos de auditoría requieren una estimación de confianza del 95% de la proporción poblacional de las facturas de ventas con errores dentro de  $\pm 0.07$ . Los resultados del mes anterior indican que la proporción más grande ha sido de no más de 0.15. De esta forma, usando la ecuación (8.5) de la página 256 y  $e = 0.07$ ,  $\pi = 0.15$  y  $Z = 1.96$  para el 95% de confianza,

$$\begin{aligned} n &= \frac{Z^2 \pi(1 - \pi)}{e^2} \\ &= \frac{(1.96)^2 (0.15)(0.85)}{(0.07)^2} \\ &= 99.96 \end{aligned}$$

Como la regla general es redondear el tamaño de la muestra hacia arriba hasta el siguiente entero para satisfacer ligeramente de más el criterio, se necesita un tamaño de muestra de 100. Entonces, el tamaño de la muestra necesario para satisfacer los requerimientos de la empresa con base en la estimación de la proporción, el nivel de confianza deseado y el error de muestreo es igual al tamaño de la muestra tomado en la página 251. El intervalo de confianza real es más reducido que el que se requiere puesto que la proporción es 0.10, mientras que se usó 0.15 para  $\pi$  en la ecuación (8.5). La figura 8.13 muestra una hoja de trabajo de Excel.

**FIGURA 8.13**

Hoja de trabajo de Excel para determinar el tamaño de la muestra para estimar la proporción de facturas de ventas con errores de la empresa de remodelaciones Saxon.

A	B
1 For the Proportion of In-Error Sales Invoices	
2	
3 Data	
4 Estimate of True Proportion	0.15
5 Sampling Error	0.07
6 Confidence Level	95%
7	
8 Intermediate Calculations	
9 Z Value	-1.9600
10 Calculated Sample Size	99.9583
11	
12 Result	
13 Sample Size Needed	100

=NORMSINV((1 - B6)/2)  
=(B9^2 \* B4 \* (1 - B4))/B5^2  
=ROUNDUP(B10, 0)

El ejemplo 8.6 proporciona una segunda aplicación para determinar el tamaño de la muestra para la estimación de la proporción poblacional.

### EJEMPLO 8.6

#### DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA UNA PROPORCIÓN DE LA POBLACIÓN

Usted desea tener un 90% de confianza al estimar la proporción de oficinistas que responden su correo electrónico dentro de una hora hasta dentro  $\pm 0.05$ . Como no se ha realizado un estudio previo, no existe información disponible de datos pasados. Determine el tamaño de la muestra necesario.

**SOLUCIÓN** Puesto que no hay información disponible de datos pasados, suponga que  $\pi = 0.50$ . Use la ecuación (8.5) de la página 256 y  $e = 0.05$ ,  $\pi = 0.50$  y  $Z = 1.645$  para un 90% de confianza,

$$n = \frac{(1.645)^2(0.50)(0.50)}{(0.05)^2}$$

$$= 270.6$$

De manera que se necesita una muestra de 271 oficinistas para estimar la proporción de la población dentro de  $\pm 0.05$  con un 90% de confianza.

## PROBLEMAS PARA LA SECCIÓN 8.4

### Aprendizaje básico

ASISTENCIA de PH Grade

**8.34** ¿Qué tamaño de muestra se requiere si se desea tener un nivel de confianza del 95% en la estimación de la media poblacional con un error de muestreo de  $\pm 5$  y una desviación estándar de 15?

ASISTENCIA de PH Grade

**8.35** ¿Qué tamaño de muestra se requiere si se desea tener un nivel de confianza del 99% en la estimación de la media poblacional con un error de muestreo de  $\pm 20$  y una desviación estándar de 100?

ASISTENCIA de PH Grade

**8.36** ¿Qué tamaño de muestra se requiere si se desea tener un nivel de confianza del 99% en la estimación de la media poblacional con un error de muestreo de  $\pm 0.04$ ?

ASISTENCIA de PH Grade

**8.37** ¿Qué tamaño de muestra se requiere si se desea tener un nivel de confianza del 95% en la estimación de la media poblacional con un error de muestreo de  $\pm 0.02$ , mientras que existe una evidencia histórica de que la proporción poblacional es aproximadamente de 0.40?

### Aplicación de conceptos

AUTO Examen

**8.38** Se planea realizar una encuesta para determinar la media anual de los gastos médicos familiares de los empleados de una gran empresa. El gerente desea tener un nivel de confianza del 95% en que la media de la muestra esté correcta con  $\pm \$50$  de la población de media anual de gastos médicos familiares. Un estudio previo indica que la desviación estándar es aproximadamente de \$400.

- ¿Qué tan grande necesita ser el tamaño de la muestra?
- Si el gerente desea estar correcto con  $\pm \$25$ , ¿qué tamaño de muestra necesita?

**8.39** Si el gerente de una tienda de artículos para pintar desea estimar la cantidad media de pintura dentro de una lata de un galón con  $\pm 0.004$  galones con un 95% de nivel de confianza y también supone que la desviación estándar es de 0.02 galones, ¿qué tamaño de muestra necesita?

**8.40** Si el gerente de control de calidad desea estimar la media de vida de focos con  $\pm 20$  horas con un nivel de confianza del 95% y también supone que la desviación estándar del proceso es de 100 horas, ¿qué tamaño de muestra necesita?

**8.41** Si la división de inspección del departamento de pesos y medidas de un condado desea estimar la media de la cantidad de bebida gaseosa que llena botellas de 2 litros con  $\pm 0.01$  litros con un nivel de confianza del 95% y también supone que la desviación estándar es de 0.05 litros, ¿qué tamaño de muestra necesita?

ASISTENCIA de PH Grade **8.42** Un grupo de consumidores desea estimar la media de la tarifa en la boleta de electricidad para el mes de julio en los hogares de una sola familia de una gran ciudad. Con base en estudios realizados en otras ciudades, la desviación estándar supuesta es de \$25. El grupo desea estimar la media de la tarifa eléctrica en la boleta del mes de julio con  $\pm \$5$  y un nivel de confianza del 99%.

- ¿Qué tamaño de muestra necesita?
- Si desea un nivel de confianza del 99%, ¿qué tamaño de muestra necesita?

**8.43** Una agencia de publicidad que sirve a la principal estación de radio desea estimar la media de la cantidad de tiempo que la audiencia escucha diariamente la radio. A partir de estudios anteriores se ha estimado la desviación estándar en 45 minutos.

- a. ¿Qué tamaño de muestra es necesario si la agencia desea tener un nivel de confianza del 90% de estar en lo correcto con  $\pm 5$  minutos?
- b. Si se desea un nivel de confianza del 99%, ¿qué tamaño de muestra es necesario?

**8.44** Suponga que una empresa de gas utilitario desea estimar la media del tiempo de espera para la instalación del servicio con  $\pm 5$  días y un nivel de confianza del 95%. La empresa no tiene acceso a datos previos, pero considera que la desviación estándar es de aproximadamente 20 días. ¿Qué tamaño de muestra se necesita?

**8.45** El Departamento de Transporte de Estados Unidos define que el vuelo de una línea aérea llega “a tiempo” si aterriza menos de 15 minutos después del tiempo programado en el Sistema de Reservación Computarizada. Los vuelos cancelados y desviados se consideran con retraso. Un estudio de las 10 aerolíneas domésticas más grandes de Estados Unidos encontró que Southwest Airlines tiene la menor proporción de retrasos, 0.1577 (Nikos Tsikriktsis y Janelle Heineke, “The Impact of Process Variation on Customer Satisfaction: Evidence from the U.S. Domestic Airline Industry”, *Decision Sciences*, invierno de 2004, 35(1):129-142). Suponga que se le pide realizar un estudio de seguimiento para Southwest Airlines cuyo objetivo sea actualizar la proporción estimada de retrasos. ¿Qué tamaño de muestra deberá emplear para estimar la proporción poblacional con un error de

- a.  $\pm 0.06$  con un 95% de nivel de confianza?
- b.  $\pm 0.04$  con un 95% de nivel de confianza?
- c.  $\pm 0.02$  con un 95% de nivel de confianza?

**8.46** En 2001 se estimó que el 45% de los hogares en Estados Unidos compraban sus abarrotes en supermercados y el 29% en clubes al mayoreo (“68th Annual Report of the Grocery Industry”, *Progressive Grocer*, abril de 2002, 29). Considere un estudio de seguimiento enfocado en el último año calendario.

- a. ¿Qué tamaño de muestra se necesita para estimar la proporción poblacional de hogares en Estados Unidos que compran sus abarrotes en supermercados con  $\pm 0.02$  y un 95% de nivel de confianza?
- b. ¿Qué tamaño de muestra se necesita para estimar la proporción poblacional de hogares en Estados Unidos que compran sus abarrotes en clubes de mayoreo con  $\pm 0.02$  y un 95% de nivel de confianza?
- c. Compare los resultados de los incisos a) y b). Explique por qué estos resultados difieren.
- d. Si usted diseñara el estudio de seguimiento, ¿usaría una muestra y preguntaría a los sujetos ambas preguntas, o seleccionaría dos muestras separadas? Explique el razonamiento tras su decisión.

**8.47** ¿Qué proporción de gente que vive en Estados Unidos utiliza Internet cuando planea salir de vacaciones? De acuerdo con la encuesta realizada por American Express, un 35% usa

Internet (A. R. Carey y K. Carter, “Solutions”, *USA Today*, 14 de enero, 2003, 1A).

- a. ¿De qué tamaño debe ser la muestra requerida para llevar a cabo un estudio de seguimiento que proporcione un 95% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.04$  de la proporción poblacional?
- b. ¿De qué tamaño debe ser la muestra requerida para llevar a cabo un estudio de seguimiento que proporcione un 99% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.04$  de la proporción poblacional?
- c. ¿De qué tamaño debe ser la muestra requerida para llevar a cabo un estudio de seguimiento que proporcione un 95% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.02$  de la proporción poblacional?
- d. ¿De qué tamaño debe ser la muestra requerida para llevar a cabo un estudio de seguimiento que proporcione un 99% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.02$  de la proporción poblacional?
- e. Discuta los efectos del cambio del nivel de confianza deseado y el nivel de error de muestreo aceptable en los requerimientos del tamaño de la muestra.

**8.48** ¿Ha tenido una presentación de negocios que haya sido interrumpida por el repicar de un teléfono celular? En una encuesta realizada con 326 hombres y mujeres de negocios, 303 respondieron “sí” y sólo 23 respondieron “no” (“You Say”, *Presentations: Technology and Techniques for Effective Communication*, enero de 2003, 18).

- a. Construya un intervalo de confianza para la proporción poblacional de hombres y mujeres de negocios que han tenido presentaciones interrumpidas por teléfonos celulares.
- b. Interprete el intervalo construido en el inciso a).
- c. ¿De qué tamaño debe ser la muestra requerida para llevar a cabo un estudio de seguimiento que proporcione un 95% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.04$  de la proporción poblacional?
- d. ¿De qué tamaño debe ser la muestra requerida para realizar un estudio de seguimiento que proporcione un 99% de nivel de confianza en que la estimación puntual será correcta con  $\pm 0.04$  de la proporción poblacional?

**8.49** Un estudio efectuado por la Reserva Federal reportó que el 52% de 4,449 familias entrevistadas en 2001 poseían acciones, directamente o a través de fondos de inversión (Barbara Hagenbaugh, “Nation’s Wealth Disparity Widens”, *USA Today*, 22 de enero, 2003, 1A).

- a. Construya un intervalo de confianza del 95% para la proporción poblacional de familias que tenían acciones en 2001.
- b. Interprete el intervalo construido en el inciso a).
- c. Para realizar un estudio de seguimiento que estime la proporción poblacional de familias que actualmente tienen acciones con  $\pm 0.01$  con un 95% de confianza, ¿a cuántas familias deberá entrevistar?



## 8.5 ESTIMACIÓN DEL INTERVALO DE CONFIANZA Y CONSIDERACIONES ÉTICAS

Pueden surgir diversos problemas éticos relacionados con la selección de muestras y con las inferencias que los acompañan. El principal problema ético se relaciona con el hecho de que si las estimaciones del intervalo se proporcionan o no junto con los estadísticos muestrales. Mencionar un estadístico muestral sin indicar al mismo tiempo los límites del intervalo de confianza (generalmente establecidos en un 95%), el tamaño de la muestra usado y una interpretación del significado del intervalo de confianza en términos que un lector pueda comprender es un hecho que genera problemas éticos. La falta al incluir la estimación del intervalo de confianza podría llevar de forma engañosa al usuario de los resultados a pensar que la estimación puntual es todo lo que se necesita para predecir las características de la población con precisión. Por eso es importante que usted indique la estimación del intervalo en un lugar destacado de cualquier comunicación escrita, junto con una explicación simple acerca del significado del intervalo de confianza. Además, deberá indicar con claridad el tamaño de la muestra.

Una de las áreas más comunes en las que surgen problemas éticos concernientes a la estimación es en la publicación de los resultados de los sondeos políticos. A menudo, los resultados de los sondeos se resaltan poniéndolos en la primera plana de un periódico y el error de muestreo implicado junto con la metodología usada se imprime en la página donde el artículo continúa, a menudo a la mitad del periódico. Para asegurar una presentación ética de los resultados estadísticos, los niveles de confianza, el tamaño de la muestra y los límites del intervalo deben estar disponibles para todas las encuestas y demás estudios estadísticos.

## RESUMEN

En este capítulo se explicó la estimación de los intervalos de confianza para las características de una población, así como la determinación del tamaño de la muestra. Usted aprendió cómo un contador de la empresa de remodelaciones Saxon puede usar los datos de una auditoría para estimar parámetros poblaciona-

les importantes, tales como la media de la cantidad en dólares de las facturas y la proporción de embarques que contienen errores. La tabla 8.3 proporciona una lista de los temas que abarcó este capítulo.

**TABLA 8.3**

Resumen de temas en el capítulo 8.

Tipo de datos		
Tipo de análisis	Numéricos	Categóricos
Intervalo de confianza para un parámetro poblacional	Estimación del intervalo de confianza para una media (secciones 8.1 y 8.2)	Estimación del intervalo de confianza para una proporción (sección 8.3)

Para determinar qué ecuación es conveniente emplear en una situación específica, usted necesita hacerse varias preguntas:

- ¿Está desarrollando un intervalo de confianza o está determinando el tamaño de una muestra?
- ¿Tiene una variable numérica o una variable categórica?

- Si tiene una variable numérica, ¿conoce la desviación estándar de la población? Si es así, utilice la distribución normal. Si no, la distribución *t*.

Los siguientes capítulos desarrollan el método de prueba de hipótesis para tomar decisiones acerca de los parámetros poblacionales.

## FÓRMULAS IMPORTANTES

### Intervalo de confianza para la media ( $\sigma$ conocida)

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} \quad (8.1)$$

$$\left| \begin{array}{l} \text{o} \\ \bar{X} - Z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z \frac{\sigma}{\sqrt{n}} \end{array} \right.$$

**Intervalo de confianza para la media ( $\sigma$  desconocida)**

$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} \quad (8.2)$$

o

$$\bar{X} - t_{n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} \frac{S}{\sqrt{n}}$$

**Estimación del intervalo de confianza para la proporción**

$$p \pm Z \sqrt{\frac{p(1-p)}{n}} \quad (8.3)$$

o

$$p - Z \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + Z \sqrt{\frac{p(1-p)}{n}}$$

**Determinación del tamaño de la muestra para la media**

$$n = \frac{Z^2 \sigma^2}{e^2} \quad (8.4)$$

**Determinación del tamaño de la muestra para la proporción**

$$n = \frac{Z^2 \pi(1-\pi)}{e^2} \quad (8.5)$$

**CONCEPTOS CLAVE**

distribución  $t$  Student 244  
error de muestreo 254

estimación del intervalo de confianza 238  
estimación puntual 238

grados de libertad 244  
nivel de confianza 241  
valor crítico 241

**PROBLEMAS DE REPASO****Revisión de su comprensión**

**8.50** ¿Por qué nunca podemos realmente tener un nivel de confianza del 100% en la estimación correcta de las características de la población de interés?

**8.51** ¿Cuándo se usa la distribución  $t$  al construir la estimación del intervalo de confianza para la media?

**8.52** ¿Por qué es cierto que para un determinado tamaño de muestra  $n$  se logra un incremento en la confianza al ampliar (y, por lo tanto, hacer menos preciso) el intervalo de confianza?

**8.53** ¿Cuándo se usa la distribución normal para construir una estimación de confianza para la media?

**8.54** Para un nivel de confianza dado, ¿por qué es el valor crítico de  $t$  menor para 80 grados de libertad que para 20 grados de libertad?

**8.55** ¿Por qué es menor el tamaño de muestra necesario para determinar la proporción cuando la proporción poblacional es 0.20 que cuando es 0.50?

**Aplicación de conceptos**

**Puede resolver los problemas 8.56 a 8.70 con o sin Excel, Minitab o SPSS. Debe usar Excel, Minitab o SPSS para resolver los problemas 8.71 a 8.74.**

**8.56** La revista *Redbook* realizó una encuesta a través de su sitio Web (Stacy Kravetz, "Work Week", *The Wall Street Journal*, 13 de abril, 1999, A1). A la gente que visitaba el sitio Web se le

dio la oportunidad de llenar una encuesta en pantalla. Un total de 665 mujeres respondieron a la pregunta de si preferían una semana laboral de 4 días o un aumento de sueldo del 20%. En total, 412 de quienes respondieron prefirieron la semana laboral de cuatro días.

- Defina la población de la que se extrajo esta muestra.
- ¿Es ésta una muestra aleatoria de la población?
- ¿Es éste un estudio válido?
- Describa cómo diseñaría un estudio estadísticamente válido para investigar la proporción de suscriptores de *Redbook* que preferirían una semana laboral de cuatro días en lugar de un aumento de sueldo del 20%. Utilice esa información para determinar el tamaño de muestra necesario para estimar la proporción poblacional con  $\pm 0.02$  con un 95% de nivel de confianza.

**8.57** En la actualidad, las empresas pasan más tiempo analizando a los aspirantes que antes. Un resultado de este proceso de estudio más exhaustivo es que las compañías encuentran que muchos aspirantes exageran la verdad en sus solicitudes. Un estudio realizado por Automatic Data Processing encontró incongruencias entre el aspirante y sus empleadores anteriores en el 44% de las solicitudes analizadas. Por desgracia, el artículo no reveló el tamaño de la muestra usada en el estudio (Stephanie Armour, "Security Checks Worry Workers", *USA Today*, 19 de junio, 2002, B1).

- Suponga que se analizaron 500 solicitudes. Construya una estimación del intervalo de confianza del 95% para la proporción poblacional de las solicitudes que contienen imprecisiones con respecto a los empleadores anteriores.

- b. Con base en el inciso a), ¿es correcto concluir que menos de la mitad de las solicitudes contienen imprecisiones con respecto a los empleadores anteriores?
- c. Suponga que el estudio usó un tamaño de muestra de 200 solicitudes. Construya una estimación del intervalo de confianza del 95% de la proporción poblacional de las solicitudes que contienen imprecisiones con respecto a los empleadores anteriores.
- d. Con base en el inciso c), ¿sería correcto concluir que menos de la mitad de las solicitudes contienen imprecisiones con respecto a los empleadores anteriores?
- e. Explique el efecto del tamaño de la muestra en sus respuestas a los incisos a) hasta d).

**8.58** Una revista de gerencia de ventas y marketing realizó una encuesta acerca de vendedores que mienten en sus reportes de gastos y que incurren en otras conductas no éticas (D. Haralson y Q. Tian, “Cheating Hearts”, *USA Today*, 15 de febrero, 2001, 1A). En la encuesta, los gerentes sorprendieron a los vendedores mintiendo en sus reportes de gastos el 58% del tiempo, laborando en un segundo trabajo durante su horario con la empresa el 50% del tiempo, citando un “bar nudista” como si fuera un restaurante en su reporte de gastos el 22% del tiempo y dando una comisión a un cliente el 19% del tiempo. Suponga que la encuesta se basa en una muestra de 200 gerentes. Construya una estimación del intervalo de confianza del 95% para la proporción poblacional de gerentes que han sorprendido a vendedores:

- a. mintiendo en un reporte de gastos.
- b. laborando en un segundo trabajo durante el horario de la empresa.
- c. citando un “bar nudista” como si fuera un restaurante en el reporte de gastos.
- d. dando comisión a un cliente.
- e. Utilice la información proporcionada en este problema para determinar el tamaño de muestra necesario para estimar la proporción de gerentes que han sorprendido a vendedores laborando en un segundo trabajo durante el horario de la empresa con  $\pm 0.02$  y un nivel de confianza del 95%.

**8.59** Los hoteles Starwood efectuaron una encuesta con 401 altos directivos que juegan golf (Del Jones, “Many CEOs Bend the Rules (of Golf)”, *USA Today*, 26 de junio, 2002). Entre los resultados obtenidos están los siguientes:

- 329 hacen trampa en el golf.
- 329 odian a aquellos que hacen trampa en el golf.
- 289 creen que el comportamiento en los negocios y el golf es paralelo.
- 80 dejarían ganar a un cliente para conseguir un negocio.
- 40 dirían que están enfermos para no ir a trabajar e irían a jugar golf.

Construya estimaciones de intervalos de confianza del 95% para cada una de estas preguntas. ¿Qué conclusiones se desprenden a partir de las actitudes de los directores generales hacia el golf a partir de estos resultados?

**8.60** El investigador de mercado de una empresa de productos electrónicos desea estudiar los hábitos de ver televisión de los residentes de un área en particular. Se selecciona una muestra aleatoria de 40 personas que respondieron, y a cada una se le instruye para mantener un registro detallado de todo lo que ve

en televisión durante una semana específica. Los resultados son los siguientes:

- Tiempo que pasan viendo televisión por semana:  $\bar{X} = 15.3$  horas,  $S = 3.8$  horas.
- 27 personas de quienes respondieron vieron el noticiero nocturno por lo menos 3 veces en la semana.
- a. Construya una estimación del intervalo de confianza del 95% para la media de la cantidad de tiempo que se pasó viendo televisión por semana en esta ciudad.
- b. Construya una estimación del intervalo de confianza del 95% para la proporción poblacional que ve el noticiero nocturno al menos 3 veces por semana.

Si el investigador de mercado deseara realizar otra encuesta en una ciudad diferente, responda a estas preguntas:

- c. ¿Qué tamaño de muestra se requiere para tener un nivel de confianza del 95% al estimar la media poblacional con  $\pm 2$  horas y suponer que la desviación estándar de la población es igual a 5 horas?
- d. ¿Qué tamaño de muestra se requiere para tener un nivel de confianza del 95% al estimar la proporción poblacional con  $\pm 0.035$  que ve las noticias por la noche al menos 3 veces por semana si no disponemos de ninguna estimación previa?
- e. Con base en los incisos c) y d), ¿qué tamaño de muestra debería seleccionar el investigador de mercado si se ha realizado una sola encuesta?

**8.61** El asesor inmobiliario del gobierno de un condado desea estudiar las diferentes características de las casas unifamiliares en ese condado. Una muestra aleatoria de 70 casas revela lo siguiente:

- Área de la casa con calefacción (en pies cuadrados)  $\bar{X} = 1,759$ ,  $S = 380$ .
- 42 casas tienen una central de aire acondicionado.
- a. Construya una estimación del intervalo de confianza del 99% de la media poblacional del área de la casa con calefacción.
- b. Construya una estimación del intervalo de la proporción poblacional de casas que tiene una central de aire acondicionado.

**8.62** El director de personal de una gran corporación desea estudiar el ausentismo entre los oficinistas en la central de la compañía durante el año. Una muestra aleatoria de 25 oficinistas revela lo siguiente:

- Ausentismo:  $\bar{X} = 9.7$  días,  $S = 4.0$  días.
- 12 oficinistas se ausentaron por más de 10 días.
- a. Construya una estimación del intervalo de confianza del 95% de la media del número de ausencias para los oficinistas en el último año.
- b. Construya una estimación del intervalo de confianza del 95% de la proporción poblacional de oficinistas que se ausentaron más de 10 días durante el año pasado.

Si el director de personal también deseara realizar una encuesta en una oficina filial, responda las siguientes preguntas:

- c. ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 95% al estimar la media poblacional con  $\pm 1.5$  días si la desviación estándar poblacional es de 4.5 días?
- d. ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% al estimar la proporción poblacional con  $\pm 0.075$  si no hay una estimación previa disponible?
- e. Con base en los incisos c) y d), ¿qué tamaño de muestra es necesario si sólo se realiza una encuesta?

**8.63** El director de investigación de mercado de la tienda departamental Dotty's desea estudiar los gastos de mujeres en cosméticos. Se diseña una encuesta para estimar la proporción de mujeres que compran sus cosméticos principalmente en la tienda departamental Dotty's, y la media de la cantidad anual que las mujeres gastan en cosméticos. Una encuesta anterior mostró que la desviación estándar de la cantidad que las mujeres gastan en cosméticos en un año es de aproximadamente \$18.

- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 99% en la estimación de la media poblacional con  $\pm 5$ ?
- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% en la estimación de la proporción poblacional con  $\pm 0.045$ ?
- Con base en los resultados de los incisos a) y b), ¿cuántos de los clientes con tarjeta de crédito de la tienda deben ser muestreados? Explique su respuesta.

**8.64** La gerente de una sucursal de una cadena de librerías a nivel nacional desea estudiar las características de los clientes de su tienda. Ella decide enfocarse en dos variables: la cantidad de dinero gastado por sus clientes y si los clientes considerarían o no comprar videos educativos relacionados con la preparación para los exámenes de graduación tales como el GMAT (Graduate Management Admission Test), GRE (Graduate Record Exam) o el LSAT (Law School Admission Test). Los resultados de una muestra de 70 clientes son los siguientes:

- Cantidad gastada:  $\bar{X} = \$28.52$ ,  $S = \$11.39$ .
  - 28 clientes afirmaron que considerarían comprar los videos educativos.
- Construya una estimación del intervalo de confianza del 95% para la media poblacional de la cantidad gastada en la librería.
  - Construya una estimación del intervalo de confianza del 90% de la proporción poblacional de clientes que considerarían comprar videos educativos.
  - Suponga que el gerente de sucursal de otra tienda de esa cadena desea realizar una encuesta similar en su tienda.
  - ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 95% al estimar la media poblacional de la cantidad gastada en su tienda con  $\pm \$2$  si se supone que la desviación estándar es de \$10?
  - ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% en la estimación de la proporción poblacional de quienes considerarían comprar los videos educativos con  $\pm 0.04$ ?
  - Con base en sus respuestas a los incisos c) y d), ¿de qué tamaño debería ser la muestra que el gerente necesita considerar?

**8.65** El gerente de sucursal de una tienda de distribución a nivel nacional de artículos para mascotas (tienda 1) desea estudiar las características de sus clientes. Particularmente decide enfocarse en dos variables: la cantidad de dinero gastado por los clientes y si los clientes poseen sólo un perro, sólo un gato o más de un perro y/o gato. Los resultados de una muestra de 70 clientes son:

- Cantidad de dinero gastado:  $\bar{X} = \$21.34$ ,  $S = \$9.22$ .
- 37 clientes sólo poseen un perro.
- 26 clientes sólo poseen un gato.

- 7 clientes poseen más de un perro y/o gato.
- Construya una estimación del intervalo de confianza del 95% de la media poblacional de la cantidad gastada en la tienda de artículos para mascotas.
  - Construya una estimación del intervalo de confianza del 90% de la proporción poblacional de clientes que sólo poseen un gato.

El gerente de sucursal de otra tienda de distribución (tienda 2) desea llevar a cabo una encuesta similar. El gerente no tiene acceso a la información generada por el gerente de la tienda 1.

- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 95% en la estimación de la media poblacional de la cantidad gastada en su tienda con  $\pm \$1.50$  si la desviación estándar es de \$10?
- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% al estimar la proporción poblacional de los clientes que sólo poseen un gato con  $\pm 0.045$ ?
- Con base en sus respuestas a los incisos c) y d), ¿de qué tamaño debería ser la muestra que el gerente necesita considerar?

**8.66** El dueño de un restaurante que sirve comida continental desea estudiar las características de sus clientes. Decide enfocarse en dos variables: la cantidad de dinero gastado por los clientes y si los clientes ordenan postres. Los resultados de una muestra de 60 clientes son los siguientes:

- Cantidad gastada  $\bar{X} = \$38.54$ ,  $S = \$7.26$ .
  - 18 clientes ordenaron postre.
- Construya una estimación del intervalo de confianza del 95% de la media poblacional de la cantidad gastada por cliente en el restaurante.
  - Construya una estimación del intervalo de confianza del 90% de la proporción poblacional de los clientes que ordenaron postre.

El dueño de un restaurante competidor desea realizar una encuesta similar en su restaurante. Este propietario no tiene acceso a la información del dueño del primer restaurante.

- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% en la estimación de la proporción poblacional de la media de la cantidad gastada en su restaurante con  $\pm \$1.50$ , suponiendo que la desviación estándar es \$8?
- ¿Qué tamaño de muestra se necesita para tener un nivel de confianza del 90% en la estimación de la proporción poblacional de clientes que ordenan postre con  $\pm 0.04$ ?
- Con base en sus respuestas a los incisos c) y d), ¿de qué tamaño debería ser la muestra que tome el dueño del restaurante?

**8.67** El fabricante de "Ice Melt" afirma que su producto derretirá la nieve y el hielo a temperaturas tan bajas como  $0^{\circ}$  Fahrenheit. El representante de una gran cadena de ferreterías está interesado en probar esta afirmación. La cadena de ferreterías compra un gran cargamento de bolsas de cinco libras para distribuir. El representante desea saber con un nivel de confianza del 95%, con  $\pm 0.05$ , qué proporción de bolsas de Ice Melt dan los resultados que asegura el fabricante.

- ¿Cuántas bolsas necesita probar el representante? ¿Qué posición debe hacerse en relación con la proporción poblacional? (A esto se le llama *prueba destructiva*; esto es, el producto probado es destruido por la prueba y entonces no es posible venderlo.) El representante prueba 50 bolsas y 42 dan los resultados que se afirmaba.

- b. Construya una estimación del intervalo de confianza del 95% para la proporción poblacional que da el resultado que se afirmaba.
- c. ¿Cómo puede el representante usar los resultados del inciso b) para determinar si debe vender el producto Ice Melt?

**8.68** Una empresa fabrica bastidores de acero para equipo electrónico. El componente principal del bastidor es un canalón de acero hecho de acero enrollado de calibre 14. Se fabrica usando una prensa que golpea con una fuerza de 250 toneladas y una operación de limpieza hacia abajo que coloca dos formas de 90 grados en el acero plano para hacer el canalón. La distancia de un lado a otro del canalón es de especial importancia por las aplicaciones a prueba de clima en la intemperie. Los datos provienen de una muestra de 49 canalones y son los siguientes:

**TROUGH**

*Amplitud del canalón (en pulgadas)*

8.312 8.343 8.317 8.383 8.348 8.410 8.351 8.373 8.481 8.422  
 8.476 8.382 8.484 8.403 8.414 8.419 8.385 8.465 8.498 8.447  
 8.436 8.413 8.489 8.414 8.481 8.415 8.479 8.429 8.458 8.462  
 8.460 8.444 8.429 8.460 8.412 8.420 8.410 8.405 8.323 8.420  
 8.396 8.447 8.405 8.439 8.411 8.427 8.420 8.498 8.409

- a. Construya una estimación del intervalo de confianza del 95% de la media de la amplitud de los canalones.
- b. Interprete el intervalo desarrollado en el inciso a).

**8.69** Una característica de calidad de interés para el proceso de relleno de las bolsas de té es el peso del té en bolsas individuales. En este ejemplo, el peso rotulado en el paquete indica que la media de la cantidad es de 5.5 gramos de té por bolsa. Si las bolsas no están bien llenadas, surgen dos problemas. Primero, los clientes pueden no lograr una infusión de té tan fuerte como lo desean. Segundo, la empresa podría estar violando las leyes de rotular con la verdad. Por otro lado, si la media de la cantidad de té en una bolsa excede el peso rotulado, la empresa está regalando producto. Llenar una bolsa con la cantidad exacta de té es problemático por la variación de temperatura y humedad dentro de la fábrica, por las diferencias de densidad del té y por la operación de llenado de la máquina que es excesivamente rápida (aproximadamente 170 bolsas por minuto). La siguiente tabla proporciona el peso en gramos de una muestra de 50 bolsas de té producidas en una hora por una sola máquina.

**TEABAGS**

*Peso de las bolsas de té en gramos*

5.65 5.44 5.42 5.40 5.53 5.34 5.54 5.45 5.52 5.41  
 5.57 5.40 5.53 5.54 5.55 5.62 5.56 5.46 5.44 5.51  
 5.47 5.40 5.47 5.61 5.53 5.32 5.67 5.29 5.49 5.55  
 5.77 5.57 5.42 5.58 5.58 5.50 5.32 5.50 5.53 5.58  
 5.61 5.45 5.44 5.25 5.56 5.63 5.50 5.57 5.67 5.36

- a. Construya una estimación del intervalo de confianza del 99% para la media poblacional del peso de las bolsas de té.
- b. ¿Está cumpliendo la empresa con el requerimiento establecido en el rótulo frontal donde se indica que la media de la cantidad de té es de 5.5 gramos en una bolsa?

**8.70** El fabricante de tablillas de asfalto “Boston” y “Vermont” ofrece a sus clientes una garantía de 20 años en la mayo-

ría de sus productos. Para determinar si una tablilla durará todo el periodo de la garantía, se realiza una prueba de aceleración de vida en la fábrica. La prueba de aceleración de vida expone las tablillas a los factores a los que estarían sujetas en el curso de una vida normal, usando un experimento en el laboratorio que dura sólo unos cuantos segundos. En esta prueba, se frota repetidamente la tablilla con un cepillo por un corto periodo de tiempo y se pesa (en gramos) la cantidad de gránulos de tablilla que se desprenden por el cepillado. Se espera que las tablillas que experimentan una baja cantidad de pérdida de gránulos duren más en un uso normal que las tablillas que experimentan una gran cantidad de pérdida de gránulos. En esta situación, una tablilla debería experimentar una pérdida de gránulos no mayor a 0.8 gramos si se espera que dure el periodo de garantía. El archivo de datos **GRANULE** contiene una muestra de 170 medidas hechas en la compañía de tablillas Boston, y 140 medidas hechas en las tablillas Vermont.

- a. Construya una estimación del intervalo de confianza del 95% de la media de la pérdida de gránulos para las tablillas Boston.
- b. Construya una estimación del intervalo de confianza del 95% de la media de la pérdida de gránulos para las tablillas Vermont.
- c. Evalúe si la suposición necesaria para los incisos a) y b) se ha infringido seriamente.
- d. Con base en los resultados de los incisos a) y b), ¿a qué conclusiones se llega respecto a la media de la pérdida de gránulos de las tablillas Boston y Vermont?

**8.71** El productor de las tablillas de asfalto “Boston” y “Vermont” sabe que el peso del producto es un factor importante en la percepción de calidad que tiene el cliente. La última etapa en la línea de ensamblaje empaca las tablillas antes de que sean colocadas en paletas de madera. Una vez que la paleta está llena (para casi todas las marcas, una paleta sostiene 16 cuadros de tablillas), es pesada y se registra la medida. Los datos del archivo **PALLET** contienen el peso (en libras) de una muestra de 368 paletas de tablillas Boston y 330 paletas de tablillas Vermont.

- a. Construya una estimación del intervalo de confianza del 95% de la media del peso para las tablillas Boston.
- b. Construya una estimación del intervalo de confianza del 95% de la media del peso para las tablillas Vermont.
- c. Evalúe si la suposición necesaria para los incisos a) y b) se ha infringido seriamente.
- d. Con base en los resultados de los incisos a) y b), ¿a qué conclusiones se llega respecto a la media del peso de las tablillas Boston y Vermont?

**8.72** Zagat publica las clasificaciones de los restaurantes en diferentes ciudades de Estados Unidos. El archivo de datos **RESTRATE** contiene la clasificación de Zagat para comida, decoración y servicio, así como el precio por persona de una muestra de 50 restaurantes localizados en la ciudad de Nueva York y 50 restaurantes localizados en Long Island, un suburbio de la ciudad de Nueva York.

*Fuente:* Zagat Survey 2002 New York City Restaurants y Zagat Survey 2001/2002 Long Island Restaurants.

Para los restaurantes de la ciudad de Nueva York y Long Island separadamente:

- Construya una estimación del intervalo de confianza del 95% para la media de la clasificación comida, para la media de la clasificación decoración, para la media de la clasificación servicio y para la media del precio por persona.
- ¿A qué conclusiones se llega acerca de los restaurantes en la ciudad de Nueva York y de Long Island a partir de los resultados en el inciso a)?

### Ejercicios de reporte escrito

- 8.73** Escriba un reporte que resuma sus conclusiones respecto a los resultados del problema 8.68 en la página 264 en relación con la anchura del canalón de acero.



### PROYECTO EN EQUIPO

- 8.74** Remítase al proyecto en equipo en la página 62. Construya todos los intervalos de confianza de las características de la población de bajo riesgo, de riesgo promedio y de alto riesgo de los fondos de inversión que sean apropiados. Incluya estas estimaciones en un reporte dirigido al vicepresidente de investigación del servicio de inversión financiera. **MUTUALFUNDS2004**

## CASO ACTUAL

### ADMINISTRACIÓN DEL SPRINGVILLE HERALD

El departamento de marketing ha estado considerando formas para incrementar tanto el número de nuevas suscripciones como la tasa de retención entre aquellos clientes que acepten una suscripción a prueba. Siguiendo la sugerencia del asistente de la gerencia, Lauren Alfonso, el personal del departamento diseñó una encuesta para ayudar a determinar las diferentes características de los lectores del periódico que no eran suscriptores de entrega a domicilio. La encuesta consistió en las siguientes 10 preguntas:

- ¿Usted o un miembro de su hogar ha comprado alguna vez el *Springville Herald*?
  - Sí
  - No

[Si la respuesta es No, concluye la entrevista.]
- ¿Recibe usted el *Springville Herald* a través del servicio de entrega a domicilio?
  - Sí
  - No

[Si la respuesta es No, pase a la pregunta 4.]
- Recibe usted el *Springville Herald*:
  - De lunes a sábado
  - Sólo el domingo
  - Diariamente

[Si la respuesta es diariamente, pase a la pregunta 9.]
- Con cuánta frecuencia compra el *Springville Herald* en el periodo de lunes a sábado?
  - Todos los días
  - Casi todos los días
  - Ocasionalmente o nunca
- ¿Qué tan a menudo compra el *Springville Herald* en domingo?
  - Cada domingo
  - 2 o 3 domingos al mes
  - No más de una vez por mes
- ¿Dónde es más probable que compre el *Springville Herald*?
  - En una tienda al menudeo
  - En un puesto de periódicos o de dulces
  - En una máquina expendedora
  - En el supermercado
  - Otro
- Consideraría suscribirse al *Springville Herald* por un periodo de prueba si se le ofreciera un descuento?
  - Sí
  - No

[Si la respuesta es No, pase a la pregunta 9.]

- Actualmente el *Springville Herald* cuesta 50 centavos de lunes a sábado y \$1.50 los domingos, para un total de \$4.50 por semana. ¿Cuánto estaría dispuesto a pagar para tener un servicio de entrega a domicilio por un periodo de prueba de 90 días?

- ¿Lee algún otro periódico que no sea el *Springville Herald*?

(1) Sí (2) No

- Como incentivo para los suscriptores de largo plazo, el periódico está considerando la posibilidad de ofrecer una tarjeta que les proporcione descuentos en restaurantes seleccionados en el área de Springville para todos los suscriptores que paguen seis meses de entrega a domicilio por adelantado. ¿Le gustaría obtener esta tarjeta de acuerdo con los términos de esta oferta?

(1) Sí (2) No

El grupo estuvo de acuerdo en utilizar un método de marcado aleatorio para encuestar a 500 hogares locales por teléfono. Usando este método, los últimos cuatro dígitos de un número telefónico son seleccionados aleatoriamente para que vayan con un código de área y cambien (los primeros 6 dígitos de un número telefónico de 10 dígitos). Sólo los pares de códigos de área y cambios que correspondían al área de Springville se utilizaron para esta encuesta.

De los 500 hogares seleccionados, 94 se rehusaron a participar, no se logró establecer comunicación con ellos después de repetidos intentos, o tenían números telefónicos que no estaban en servicio. El resumen de los resultados es el siguiente:

#### Hogares que compran el *Springville Herald*

	Frecuencia
Sí	352
No	54

#### Hogares con entrega a domicilio

	Frecuencia
Sí	136
No	216

<b>Tipo de suscripción</b>	<b>Frecuencia</b>
Lunes a sábado	18
Domingos únicamente	25
Toda la semana	93
<b>Comportamiento de compra de los no suscriptores para las ediciones de lunes a sábado</b>	
	<b>Frecuencia</b>
Todos los días	78
Casi todos los días	95
Ocasionalmente o nunca	43
<b>Comportamiento de compra de los no suscriptores para las ediciones del domingo</b>	
	<b>Frecuencia</b>
Todos los domingos	138
2 o 3 domingos al mes	54
No más de una vez al mes	24
<b>Compra en locación de los no suscriptores</b>	
	<b>Frecuencia</b>
Tienda de menudeo	74
Puesto de periódicos/tienda de dulces	95
Máquina expendedora	21
Supermercado	13
Otras locaciones	13
<b>Considerarían la suscripción de prueba si ofrecen descuento</b>	
	<b>Frecuencia</b>
Sí	46
No	170

## CASO WEB

Aplique su conocimiento sobre la estimación del intervalo de confianza en este Caso Web que proviene del Caso Web *OnCampus!* del capítulo 6.

Entre otras características, el sitio Web *OnCampus!* permite que los clientes compren en línea mercancía *OnCampus! LifeStyles*. Para manejar el proceso de pago, la gerencia de *OnCampus!* ha contratado las siguientes firmas:

- PayAFriend (PAF): Un sistema de pagos en línea en el que los clientes y negocios tales como *OnCampus!* se registran para realizar pagos de forma segura y conveniente sin necesidad de usar tarjeta de crédito.
- Continental Banking Company (Conbanco): Un proveedor de servicios de procesamiento que permite a los clientes de *OnCampus!* pagar mercancía con tarjetas de crédito reconocidas a nivel nacional expedidas por una institución financiera.

Para reducir costos, la gerencia considera eliminar uno de estos dos sistemas de pago. Sin embargo, Virginia Duffy, del departamento de ventas, considera que los clientes usan las

<b>Precio que estaría dispuesto a pagar por semana (en dólares) por una suscripción de prueba de 90 días con entrega a domicilio</b>	<b>SH8</b>
\$4.15	3.60
3.60	4.10
4.10	3.60
3.60	3.60
3.60	4.40
4.40	3.15
3.15	4.00
4.00	3.75
3.75	4.00
3.25	3.30
3.30	3.75
3.75	3.65
3.65	4.00
4.00	4.10
4.10	3.90
3.90	3.50
3.50	3.75
3.75	3.30
3.30	3.85
3.85	3.20
3.20	4.40
4.40	3.80
3.80	3.40
3.40	3.50
3.50	2.85
2.85	3.75
3.75	3.90
3.90	

<b>Lee un periódico que no es el <i>Springville Herald</i></b>	<b>Frecuencia</b>
Sí	138
No	214

<b>Pagaría con anticipación 6 meses para recibir tarjeta de descuento en restaurantes</b>	<b>Frecuencia</b>
Sí	66
No	286

## EJERCICIOS

- SH8.1 Algunos miembros del departamento de marketing están preocupados acerca del método de marcado aleatorio de dígitos usado para recolectar las respuestas de la encuesta. Prepare un memorando que examine los siguientes temas:
- Las ventajas y desventajas de usar el método de marcado aleatorio de dígitos.
  - Posibles métodos alternativos para llevar a cabo la encuesta y sus ventajas y desventajas.
- SH8.2 Analice los resultados de la encuesta de Springville en hogares. Escriba un reporte que describa las implicaciones de mercado en los resultados de la encuesta del *Springville Herald*.

dos formas de pago de forma desigual y que muestran diferentes comportamientos de compra al usar una u otra forma de pago. Por esa razón, ella quisiera determinar primero:

- La proporción de clientes que usan PAF y la proporción de clientes que usan tarjeta de crédito para pagar sus compras.
  - La media de la cantidad de compra al usar PAF y la media de la cantidad de compra cuando se usa tarjeta de crédito.
- Ayude a Virginia Duffy preparando un análisis apropiado basado en una muestra aleatoria de 50 transacciones que ella ha preparado y colocado en un archivo interno en el sitio Web *OnCampus!* [www.prenhall.com/Springville/OnCampus\\_PyMtSample.htm](http://www.prenhall.com/Springville/OnCampus_PyMtSample.htm). Resuma sus descubrimientos y determine si las conjeturas de Virginia Duffy acerca de los comportamientos de los clientes de *OnCampus!* son correctas. Si desea que el error muestral no sea mayor que \$3, ¿la muestra elegida por Duffy es lo suficientemente grande como para desarrollar un análisis válido?

## REFERENCIAS

1. Cochran, W. G., *Sampling Techniques*, 3a. ed. (Nueva York: Wiley, 1977).
2. Fosher, R. A. y F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 5a. ed. (Edimburgo: Oliver & Boyd, 1957).
3. Kirk, R.E., ed., *Statistical Issues: A Reader for the Behavioral Sciences* (Belmont, CA: Wadsworth, 1972).
4. Larsen, R. L. y M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, 2a. ed. (Englewood Cliffs, NJ: Prentice Hall, 1986).
5. Microsoft Excel 2003 (Redmond, WA: Microsoft Corporation, 2003).
6. Minitab for Windows Version 14 (State Collge, PA: Minitab Inc., 2004).
7. Snedecor, G. W. y W. G. Cochran, *Statistical Methods*, 7<sup>a</sup>. ed. (Ames; IA: Iowa State University Press, 1980).
8. SPSS Base 12.0 Brief Guide (Upper Saddle River, NJ: Prentice Hall, 2003).

# Apéndice 8 Uso del software para los intervalos de confianza y la determinación del tamaño de la muestra

## A8.1 EXCEL

### Para la estimación del intervalo de confianza de la media ( $\sigma$ conocida)

Abra el archivo **CIE sigma known.xls**. Esta hoja de trabajo contiene las entradas para el ejemplo 8.1 en la página 242 y utiliza las funciones NORMSINV y CONFIDENCE (vea la sección G.11 para mayor información). Para adaptar esta hoja de trabajo a otros problemas, cambie los valores de la desviación estándar poblacional, la media muestral, el tamaño de la muestra y el nivel de confianza en las celdas sombreadas de las filas 4 a la 7.

O Vea la sección G.11 (Estimación del intervalo de confianza para la media,  $\sigma$  conocida) si desea que PHStat2 genere una hoja de trabajo.

### Para la estimación del intervalo de confianza de la media ( $\sigma$ desconocida)

Abra el archivo **CIE sigma unknown.xls**, mostrado en la figura 8.6 de la página 246. La hoja de trabajo usa la función TINV para determinar el valor crítico de la distribución  $t$  (vea la sección G.12 para más información).

Para adaptar esta hoja de trabajo a otros problemas, cambie los estadísticos de la muestra y los valores del nivel de confianza en las celdas sombreadas de las filas 4 a la 7.

O Vea la sección G.12 (Estimación del intervalo de confianza para la media,  $\sigma$  desconocida) si desea que PHStat2 genere una hoja de trabajo.

### Para la estimación del intervalo de confianza para la proporción

Abra el archivo **CIE Proportion.xls**, mostrado en la figura 8.10 en la página 251. La hoja de trabajo utiliza la función NORMSINV para determinar el valor  $Z$  (vea la sección G.13 para mayor información).

Para adaptar esta hoja de cálculo a otros problemas, cambie los valores del tamaño de la muestra, el número de éxitos y el nivel de confianza en las celdas sombreadas en las filas 4 a 6.

O Vea la sección G.13 (Estimación del intervalo de confianza para la proporción) si desea que PHStat2 genere una hoja de trabajo.

### Para la determinación del tamaño de la muestra para la media

Abra el archivo **Sample Size Mean.xls**, mostrado en la figura 8.12 en la página 255. La hoja de cálculo utiliza las funciones NORMSINV y ROUNDUP (vea la sección G.14 para mayor información).

Para adaptar esta hoja de trabajo a otros problemas, cambie los valores de la desviación estándar poblacional, el error de muestreo y el nivel de confianza en las celdas sombreadas de las filas 4 a la 6.

O Vea la sección G.14 (Determinación del tamaño de la muestra para la media) si desea que PHStat2 genere una hoja de trabajo.

### Para la determinación del tamaño de la muestra para la proporción

Abra el archivo **Sample Size Proportion.xls**, mostrado en la figura 8.13 en la página 257. La hoja de trabajo usa las funciones NORMSINV y ROUNDUP (vea la sección G.15 para mayor información).

Para adaptar esta hoja de trabajo a otros problemas, cambie los valores de la proporción real, el error de muestreo y el nivel de confianza en las celdas sombreadas de las filas 4 a la 6.

O Vea la sección G.15 (Determinación del tamaño de la muestra para la proporción) si desea que PHStat2 genere una hoja de trabajo.

## A8.2 MINITAB

### Uso de Minitab para la estimación del intervalo de confianza para la media

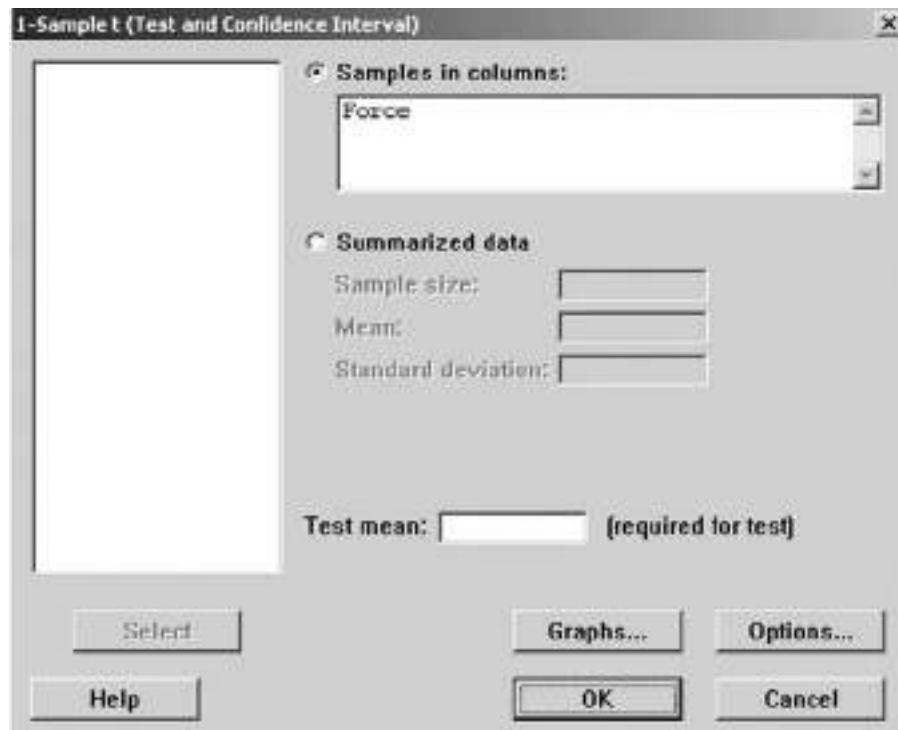
Puede usar Minitab para crear la estimación del intervalo de confianza para la media cuando la  $\sigma$  es conocida seleccionando **Stat → Basic Statistics → 1-Sample Z**. Ingrese el nombre de la variable en el cuadro de edición Samples in columns. Ingrese el valor de  $\sigma$  en el cuadro de edición Standard deviation. Seleccione el botón **Options** e ingrese el nivel de confianza en el cuadro de edición Confidence level. Dé clic en **OK**. Dé clic en este botón nuevamente.

Puede usar Minitab para crear una estimación del intervalo de confianza para la media cuando  $\sigma$  es desconocida seleccionando **Stat → Basic Statistics → 1-Sample t**. Para calcular la estimación del intervalo de confianza de la media poblacional para la fuerza para romper los aislantes del ejemplo 8.3 en la página 247, abra la hoja de trabajo **FORCE.MTW** y haga lo siguiente:

1. Seleccione **Stat → Basic Statistics → 1-Sample t**.
2. En la ventana de diálogo 1-Sample t (Prueba e intervalo de confianza) (vea la figura A8.1), seleccione la opción **Samples in columns**. Ingrese **C1** o **Force** en el cuadro de edición Samples in columns. (Si ya ha resumido los datos, seleccione el botón con la opción **Summarized data** e ingrese el tamaño de la muestra, la media muestral y la desviación estándar de la muestra en las ventanas de edición).



**FIGURA A8.2** Ventana de diálogo Opción Minitab 1-Sample t.

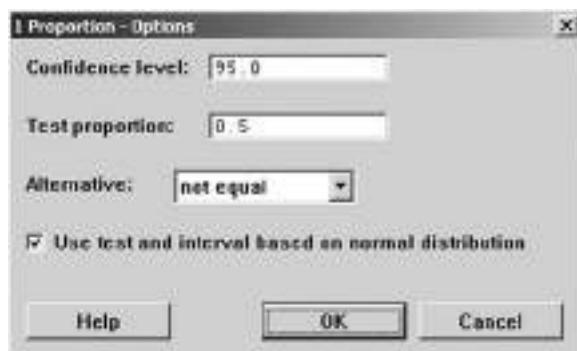


**FIGURA A8.1** Ventana de diálogo Minitab 1-Sample t.

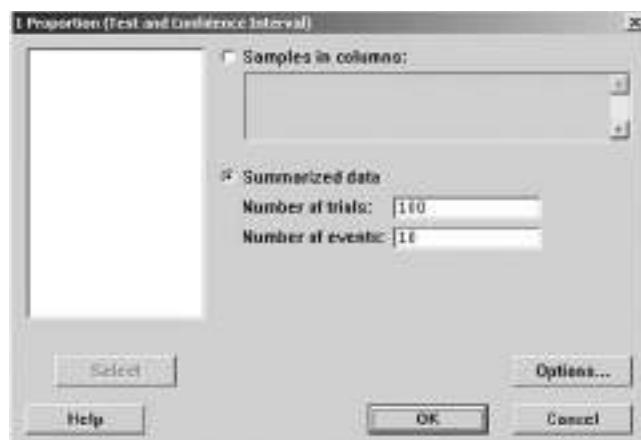
3. Seleccione **Options** (vea figura A8.2). En la ventana de diálogo 1-Sample t-Options, ingrese **95.0** en el cuadro de edición de Confidence level. Dé clic en **OK**.
4. Dé clic en **OK** de nuevo.

### Uso de Minitab para la estimación del intervalo de confianza de la proporción

Puede usar Minitab para calcular la estimación del intervalo de confianza para la proporción. Para calcular la estimación del in-



**FIGURA A8.4** Ventana de diálogo Minitab 1 Proportion-Options.



**FIGURA A8.3** Ventana de diálogo Minitab 1 Proportion.

ervalo de confianza para la proporción poblacional del problema en la sección 8.3 de la empresa de remodelaciones Saxon,

1. Seleccione **Stat → Basic Statistics → 1 Proportion**.
2. En la ventana de diálogo 1 Proportion (vea la figura A8.3), seleccione la opción **Summarized data**. Ingrese **100** en el cuadro de edición Number of trials. Ingrese **10** en el cuadro de edición Number of events.
3. Dé clic en el botón **Options**. En la ventana de diálogo Proportion-Options (vea la figura A8.4), ingrese **95** en el cuadro de edición Confidence level. Seleccione **Use test and interval based on normal distribution** en el cuadro de exploración. Dé clic en el botón **OK** para regresar a la ventana de diálogo 1 proportion.
4. Dé clic en **OK**.



# CAPÍTULO 9

## Fundamentos de la prueba de hipótesis: pruebas de una muestra

**USO DE LA ESTADÍSTICA:** Una visita más a la Oxford Cereal Company

### 9.1 METODOLOGÍA DE LA PRUEBA DE HIPÓTESIS

Hipótesis nula y alternativa  
Valor crítico del estadístico de prueba  
Regiones de rechazo y aceptación  
Riesgos de la toma de decisiones al utilizar la metodología de la prueba de hipótesis

### 9.2 PRUEBA Z DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ CONOCIDA)

Método del valor crítico para la prueba de hipótesis  
Método del valor- $p$  para la prueba de hipótesis  
Conexión entre la estimación del intervalo de confianza y la prueba de hipótesis

### 9.3 PRUEBAS DE UNA COLA

Método del valor crítico  
Método del valor- $p$

### 9.4 PRUEBA t DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ DESCONOCIDA)

Método del valor crítico  
Método del valor- $p$   
Revisión de suposiciones

### 9.5 PRUEBA Z DE HIPÓTESIS PARA LA PROPORCIÓN

Método del valor crítico  
Método del valor- $p$

### 9.6 POSIBLES OBSTÁCULOS EN LAS PRUEBAS DE HIPÓTESIS Y CONSIDERACIONES ÉTICAS

### A.9 USO DE SOFTWARE PARA LAS PRUEBAS DE HIPÓTESIS CON UNA MUESTRA

A9.1 Excel  
A9.2 Minitab  
A9.3 (Tema de CD-ROM) SPSS

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá:

- Los principios básicos de las pruebas de hipótesis
- Cómo utilizar las pruebas de hipótesis para comprobar una media o proporción
- Las suposiciones de cada uno de los procedimientos de prueba de hipótesis, cómo evaluarlos, y las consecuencias si se infringen seriamente
- Cómo evitar los obstáculos implícitos en las pruebas de hipótesis
- Los asuntos éticos implicados en las pruebas de hipótesis

# USO DE LA ESTADÍSTICA



## Una visita más a la Oxford Cereal Company

Como en el escenario “Uso de la estadística” del capítulo 7, imagine que usted es el gerente operativo de la Oxford Cereal Company y responsable de vigilar la cantidad empacada en cada caja de cereal. Selecciona y pesa una muestra aleatoria de 25 cajas, con el fin de calcular la media muestral e investigar qué tanto se acercan los pesos de llenado a la especificación de 368 gramos establecida por la empresa. Esta vez debe tomar una decisión y concluir si el peso medio de llenado de todo el proceso es igual a 368 gramos (o no), con objeto de saber si el proceso de llenado requiere de ajustes. ¿Cómo puede tomar esta decisión de una manera razonada?

A diferencia del capítulo 7, donde el problema que enfrentaba el gerente de operaciones consistía en determinar si la media muestral era congruente con la media poblacional conocida, este escenario “Uso de la estadística” pregunta: ¿Cómo se logra validar la afirmación de que la media poblacional es de 368 gramos, utilizando la media muestral? Para validar tal afirmación, primero se necesita establecer la afirmación sin ambigüedades. Por ejemplo, la media poblacional es de 368 gramos. En el método inferencial conocido como prueba de hipótesis, se considera la evidencia (el estadístico de muestra) para ver si ésta respalda mejor al enunciado, llamado *hipótesis nula*, o la alternativa mutuamente excluyente que, en este caso, establece que la media poblacional no es de 368 gramos.

En este capítulo nos concentraremos en la prueba de hipótesis, otro aspecto de la inferencia estadística que, al igual que la estimación del intervalo de confianza, se basa en información de la muestra. Se desarrolla una metodología paso a paso que le permita hacer inferencias sobre un parámetro poblacional mediante el *análisis diferencial* entre los resultados observados (estadístico de la muestra) y los resultados esperados si la hipótesis subyacente es realmente cierta. Por ejemplo, ¿es el peso medio de las cajas de cereal incluidas en la muestra obtenida de la Oxford Cereal Company un valor congruente con lo que usted esperaría si la media de toda la población de cajas de cereales es de 368 gramos? ¿O puede usted inferir que la media poblacional no es igual a 368 gramos porque la media muestral es significativamente diferente a 368 gramos?

## 9.1 METODOLOGÍA DE LA PRUEBA DE HIPÓTESIS

### Hipótesis nula y alternativa

La **prueba de hipótesis** suele comenzar con alguna teoría, afirmación, o aseveración sobre un parámetro específico de una población. Por ejemplo, su hipótesis inicial sobre el ejercicio del cereal es que el proceso está funcionando adecuadamente, lo que significa que el promedio de llenado es de 368 gramos, y que no es necesario emprender acciones correctivas.

La hipótesis de que el parámetro poblacional es igual a la especificación de la empresa se denomina hipótesis nula. Una **hipótesis nula** siempre es una de *status quo*, y se denota mediante el símbolo  $H_0$ . Aquí, la hipótesis nula expresa que el proceso de llenado está funcionando de manera adecuada y, por lo tanto, el llenado medio concuerda con la especificación de 368 gramos. Esto se establece como:

$$H_0: \mu = 368$$

A pesar de que sólo se cuenta con información de la muestra, la hipótesis nula se escribe en términos de la población. Recuerde, usted se concentra en la población de todas las cajas de cereal. El estadístico de muestra se utiliza para hacer inferencias sobre todo el proceso de llenado. Una inferencia

podría ser que los resultados observados a partir de los datos de la muestra indican que la hipótesis es falsa. Si se considera que la hipótesis nula es falsa, entonces habrá otra afirmación que debe ser cierta.

Siempre que se especifica una hipótesis nula, también se determina una hipótesis alternativa que debe ser cierta si la hipótesis nula es falsa. La **hipótesis alternativa**  $H_1$  es opuesta a la hipótesis nula  $H_0$ . En el ejemplo del cereal, esto se determina como:

$$H_1: \mu \neq 368$$

La hipótesis alternativa representa la conclusión obtenida al rechazar la hipótesis nula. Cuando a partir de la información de la muestra existe suficiente evidencia de que es falsa, se rechaza la hipótesis nula. En el ejemplo del cereal, si los pesos de las cajas muestreadas están lo bastante arriba o abajo de la media de 368 gramos especificada por la empresa, usted rechaza la hipótesis nula en favor de la hipótesis alternativa de que el llenado medio es distinto de 368 gramos. Detiene la producción y emprende la acción necesaria para corregir el problema. Si no se rechaza la hipótesis nula, entonces debe continuar confiando en el *status quo* de que el proceso funciona correctamente y, por lo tanto, no es necesaria una acción correctiva. En este segundo escenario, usted no ha demostrado que el proceso está funcionando de forma correcta. Más bien, no ha podido demostrar que está funcionando de manera incorrecta y, por lo tanto, continúa confiando (aunque sin pruebas) en la hipótesis nula.

En la metodología de prueba de hipótesis, la hipótesis nula se rechaza cuando la evidencia muestral sugiere que es más probable que ésta sea cierta que la hipótesis alternativa. Sin embargo, el no poder rechazar la hipótesis nula no comprueba que sea cierta. Usted nunca podrá demostrar que la hipótesis nula es correcta, porque la decisión se basa sólo en información de la muestra, no en toda la población. Por lo tanto, si no rechaza la hipótesis nula, sólo puede concluir que no existe evidencia suficiente para garantizar su rechazo. Los siguientes puntos fundamentales resumen las hipótesis nula y alternativa:

- La hipótesis nula  $H_0$  representa al *status quo* o creencia actual en una situación.
- La hipótesis alternativa  $H_1$  es lo opuesto a la hipótesis nula y representa una afirmación de investigación o inferencia específica que quisiera demostrar.
- Si usted rechaza la hipótesis nula, tiene una prueba estadística de que la hipótesis alternativa es correcta.
- Si usted no rechaza la hipótesis nula, entonces no ha podido demostrar la hipótesis alternativa. Sin embargo, el no poder demostrar la hipótesis alternativa no quiere decir que haya demostrado la hipótesis nula.
- La hipótesis nula  $H_0$  siempre se refiere a un valor específico del parámetro poblacional (como  $\mu$ ), no a un estadístico de muestra (como  $\bar{X}$ ).
- El enunciado que describe la hipótesis nula siempre contiene un signo de igual relacionado con el valor específico del parámetro poblacional (por ejemplo,  $H_0: \mu = 368$  gramos).
- El enunciado que describe la hipótesis alternativa nunca contiene un signo de igual relacionado con el valor específico del parámetro poblacional (por ejemplo,  $H_1: \mu \neq 368$  gramos).

## EJEMPLO 9.1

### HIPÓTESIS NULA Y ALTERNATIVA

Usted es gerente de un restaurante de comida rápida. Quiere determinar si el tiempo de espera al pedir una orden se ha modificado durante el último mes con respecto a su valor de media poblacional previo de 4.5 minutos. Determine las hipótesis nula y alternativa.

**SOLUCIÓN** La hipótesis nula dice que la media poblacional no ha cambiado con respecto a su valor previo de 4.5 minutos. Esto se establece como:

$$H_0: \mu = 4.5$$

La hipótesis alternativa es opuesta a la hipótesis nula. Puesto que la hipótesis nula dice que la media poblacional es 4.5 minutos, la hipótesis alternativa dice que la media poblacional no es 4.5 minutos. Esto se establece como:

$$H_1: \mu \neq 4.5$$

## Valor crítico del estadístico de prueba

La lógica subyacente a la metodología de prueba de hipótesis radica en determinar qué tan probable es que la hipótesis nula sea cierta, considerando la información recabada en una muestra. En el escenario de la Oxford Cereal Company, la hipótesis nula establece que la cantidad media de cereal por caja en todo el proceso de llenado es de 368 gramos (es decir, el parámetro poblacional especificado por la empresa). Del proceso de llenado, usted selecciona una muestra de cajas, las pesa y calcula la media muestral. Este estadístico es una estimación del parámetro correspondiente (la media poblacional  $\mu$ ). Incluso si la hipótesis nula es cierta, es probable que el estadístico (la media muestral  $\bar{X}$ ) sea distinto al valor del parámetro (la media poblacional  $\mu$ ), por la variación causada por el muestreo. Sin embargo, si la hipótesis nula es cierta, cabe esperar que el estadístico de muestra sea cercano al parámetro poblacional. Si el estadístico de muestra es cercano al parámetro poblacional, usted no cuenta con evidencia suficiente para rechazar la hipótesis nula. Por ejemplo, si la media poblacional es 367.9, usted concluiría que la media poblacional no cambió (es decir,  $\mu = 368$ ), porque la media muestral de 367.9 es muy cercana al valor establecido de 368 de la hipótesis. Por intuición, usted considera probable que pueda tener una media muestral de 367.9 a partir de una población con una media poblacional de 368.

Por otra parte, si existe una gran diferencia entre el valor del parámetro poblacional y el estadístico establecido en la hipótesis, usted concluirá que ésta última es falsa. Por ejemplo, si la media muestral es 320, usted concluiría que la media poblacional no es 368 (es decir,  $\mu \neq 368$ ), porque la media muestral es muy lejana al valor de 368 establecido en la hipótesis. En ese caso, usted concluye que es muy poco probable obtener una media muestral de 320 si la media poblacional es en realidad de 368. Por lo tanto, es más lógico concluir que la media poblacional no es igual a 368. Aquí usted rechaza la hipótesis nula.

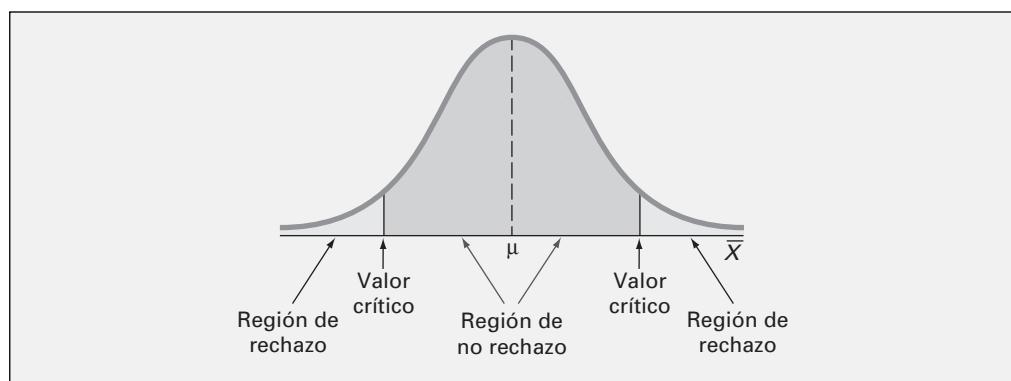
Por desgracia, el proceso de toma decisiones no siempre es tan claro. Determinar lo que es “muy cercano” y lo que es “muy diferente” es un proceso arbitrario sin definiciones claras. La metodología prueba de hipótesis brinda definiciones claras para evaluar las diferencias. Además, le permite cuantificar el proceso de toma de decisiones mediante el cálculo de la probabilidad de obtener un resultado muestral dado si la hipótesis nula resulta verdadera. Calcule esta probabilidad determinando la distribución muestral correspondiente al estadístico muestral de interés (es decir, la media muestral) y luego calcule el **estadístico de prueba** específico, con base en el resultado muestral dado. Puesto que la distribución muestral para el estadístico muestral suele seguir una distribución estadística conocida, como la distribución normal estandarizada o distribución  $t$ , éstas se utilizan como ayuda para determinar si la hipótesis nula es cierta.

## Regiones de rechazo y aceptación

La distribución muestral del estadístico de muestra se divide en dos **regiones**, una **de rechazo** (a veces llamada **región crítica**) y una **de aceptación** (vea la figura 9.1).

**FIGURA 9.1**

Regiones de rechazo y aceptación en la prueba de hipótesis.



Si el estadístico de prueba queda en la región de aceptación, no rechace la hipótesis nula. En el escenario de la Oxford Cereal Company, usted concluye que no existe evidencia suficiente de que el llenado medio poblacional sea diferente de 368 gramos. Si el estadístico de prueba queda en la región de rechazo, usted rechaza la hipótesis nula. En este caso, concluye que la media poblacional no es de 368 gramos.

La región de rechazo se compone de los valores del estadístico de prueba con muy pocas posibilidades de presentarse en caso de que la hipótesis nula sea cierta. Es más probable que tales valores se presenten si la hipótesis nula es falsa. Por lo tanto, si un valor del estadístico de prueba queda dentro de esta *región de rechazo*, la hipótesis nula se rechaza porque ese valor tiene pocas posibilidades de presentarse si la hipótesis nula es cierta.

Para tomar una decisión con respecto a la hipótesis nula, antes debe determinarse cuál es el **valor crítico** del estadístico de prueba. Ese valor separa la región de aceptación de la región de rechazo. Determinar este valor crítico depende del tamaño de la región de rechazo. El tamaño de la región de rechazo se relaciona directamente con los riesgos implícitos al utilizar sólo evidencia muestral para tomar decisiones con respecto a un parámetro poblacional.

## Riesgos de la toma de decisiones al utilizar la metodología de la prueba de hipótesis

Al utilizar un estadístico de muestra para tomar decisiones sobre el parámetro poblacional, existe el riesgo de llegar a una conclusión equivocada. Al aplicar la metodología de prueba de hipótesis, puede cometer dos tipos de error: el error tipo I y el error tipo II.

Un **error tipo I** se presenta cuando se rechaza la hipótesis nula  $H_0$  siendo cierta y no debería rechazarse. La probabilidad de que se presente un error tipo I es  $\alpha$ .

Un **error tipo II** se presenta cuando no se rechaza la hipótesis nula  $H_0$  siendo falsa y debería rechazarse. La probabilidad de que se presente un error tipo II es  $\beta$ .

En el escenario de la Oxford Cereal Company, cometería un error tipo I si concluyera que el llenado medio de la población *no* es 368 cuando *es* 368. Por su parte, cometería un error tipo II si concluyera que el llenado medio de la población *es* 368 cuando *no* es 368.

**El nivel de significancia ( $\alpha$ )** La probabilidad de cometer un error tipo I, denotado por  $\alpha$  (la letra griega *alfa* minúscula), se denomina como **nivel de significancia** del estadístico de prueba. Tradicionalmente, el error tipo I se controla al decidir el nivel de riesgo que está dispuesto a correr al rechazar la hipótesis nula, siendo ésta cierta. Como el nivel de significancia se especifica antes de realizar la prueba de hipótesis, el riesgo de cometer un error tipo I,  $\alpha$ , está directamente bajo su control. Por lo general, se seleccionan niveles de 0.01, 0.5 o 0.10. La selección de un nivel de riesgo específico para cometer un error tipo I depende del costo que implicaría cometerlo. Después de especificar el valor de  $\alpha$ , usted conoce el tamaño de la región de rechazo, porque  $\alpha$  es la probabilidad de rechazo bajo la hipótesis nula. A partir de este hecho, se determina el valor o los valores críticos que dividen las zonas de rechazo y aceptación.

**Coeficiente de confianza** El complemento de la probabilidad de un error tipo I ( $1 - \alpha$ ) se denomina coeficiente de confianza. Al multiplicarlo por 100%, el coeficiente de confianza produce el nivel de confianza estudiado al construir intervalos de confianza (vea la sección 8.1).

El **coeficiente de confianza**,  $1 - \alpha$ , expresa la probabilidad de que la hipótesis nula  $H_0$  no se rechace cuando es cierta y no debe rechazarse. El **nivel de confianza** de una prueba de hipótesis es  $(1 - \alpha) \times 100\%$ .

En términos de la metodología de prueba de hipótesis, el coeficiente de confianza representa la probabilidad de concluir que el valor del parámetro especificado en la hipótesis nula es aceptable cuando es cierta. En el escenario de la Oxford Cereal Company, el coeficiente de confianza mide la probabilidad de concluir que el llenado poblacional medio es de 368 gramos cuando en realidad es 368 gramos.

**El riesgo  $\beta$**  La probabilidad de cometer un error tipo II se denota por medio de  $\beta$  (la letra griega *beta* minúscula). A diferencia del error tipo I, el cual usted controla mediante la selección de  $\alpha$ , la posibilidad de cometer un error tipo II depende de la diferencia que existe entre los valores hipotético y real del parámetro poblacional. Como es más fácil encontrar diferencias grandes que pequeñas, si la diferencia que existe entre los valores hipotético y real del parámetro poblacional es grande,  $\beta$  es pequeña. Por ejemplo, si la media poblacional verdadera es 330 gramos, existe una pequeña posibilidad ( $\beta$ ) de que usted deduzca que la media no ha cambiado de 368. Por otra parte, si la diferencia entre los valores hipotético y real del parámetro es pequeña, la probabilidad de que cometa un error tipo II es mayor. De tal modo, si la media poblacional es en realidad de 367 gramos, usted tiene una gran probabilidad de cometer un error tipo II al concluir que la media poblacional es todavía de 368 gramos.

**La potencia de una prueba** Al complemento de la probabilidad de un error tipo II ( $1 - \beta$ ) se le denomina potencia de una prueba estadística.

**La potencia de una prueba estadística**,  $1 - \beta$ , es la probabilidad de que se rechace la hipótesis nula cuando es falsa y debería rechazarse.

En el escenario de la Oxford Cereal Company, la potencia de la prueba es la probabilidad de concluir correctamente que el llenado poblacional medio no es de 368 gramos cuando en realidad no es de 368 gramos.

**Riesgos en la toma de decisiones: un equilibrio delicado** En la tabla 9.1 se ilustran los resultados de dos decisiones posibles (no rechazar  $H_0$  o rechazar  $H_0$ ) que puede tomar en cualquier prueba de hipótesis. Dependiendo de la decisión específica, usted podría cometer uno de los dos tipos de error, o bien, obtener uno de los dos tipos de conclusión correcta.

**TABLA 9.1**

Prueba de hipótesis y toma de decisiones.

Decisión estadística	Situación real	
	$H_0$ verdadera	$H_0$ falsa
No se rechaza $H_0$	Decisión correcta $P(\text{decisión correcta}) = 1 - \alpha$	Error tipo II $P(\text{error tipo II}) = \beta$
Se rechaza $H_0$	Error tipo I $P(\text{error tipo I}) = \alpha$	Decisión correcta $P(\text{decisión correcta}) = 1 - \beta$

Una forma de reducir la probabilidad de cometer un error tipo II consiste en aumentar el tamaño de la muestra. Por lo general, las muestras grandes permiten detectar incluso diferencias pequeñas entre valores hipotéticos y parámetros poblacionales. Para un nivel de  $\alpha$  dado, aumentar el tamaño de la muestra reducirá  $\beta$  y así se incrementará el poder de la prueba para detectar que la hipótesis nula  $H_0$  es falsa. Sin embargo, siempre existe un límite para sus recursos, y esto afectará la decisión de qué tan grande debe ser la muestra tomada. De esta forma, para un tamaño dado de la muestra, usted debe considerar las comparaciones entre los dos tipos de errores posibles. Como el riesgo del error tipo I se controla directamente, éste se reducirá seleccionando un valor más pequeño para  $\alpha$ . Por ejemplo, si son considerables las consecuencias negativas asociadas con un error tipo I, puede seleccionar  $\alpha = 0.01$  en lugar de 0.05. Sin embargo, cuando  $\alpha$  se reduce,  $\beta$  aumenta, por lo que reducir el riesgo de un error tipo I tiene como resultado un aumento en el riesgo del error tipo II. Por otra parte, si desea reducir  $\beta$ , puede seleccionar un valor mayor para  $\alpha$ . Por lo tanto, si es importante tratar de evitar un error tipo II, habrá que seleccionar una  $\alpha$  de 0.5 o 0.10 en lugar de 0.01.

En el escenario de la Oxford Cereal Company, el riesgo de cometer un error tipo I implica concluir que la media ha cambiado de los 368 gramos hipotetizados, cuando en realidad eso no ha ocurrido. El riesgo de cometer un error tipo II implica concluir que la media no ha cambiado de los 368 gramos hipotetizados, cuando en realidad eso ha ocurrido. La selección de valores razonables para  $\alpha$  y  $\beta$  depende del costo inherente a cada tipo de error. Por ejemplo, si resulta muy costoso modificar el proceso de llenado de cereal, entonces usted desearía estar muy convencido de que necesita hacer un cambio antes de efectuarlo. En este caso, el riesgo de cometer un error tipo I es más importante, y usted seleccionaría una  $\alpha$  pequeña. Por otra parte, si quiere estar muy seguro de detectar cambios con respecto a la media de 368 gramos, es más importante el riesgo de cometer un error tipo II, y entonces seleccionaría un mayor nivel de  $\alpha$ .

## PROBLEMAS PARA LA SECCIÓN 9.1

### Aprendizaje básico

**ASISTENCIA de PH Grade** **9.1** ¿Para qué hipótesis utiliza el símbolo  $H_0$ ?

**ASISTENCIA de PH Grade** **9.2** ¿Para qué hipótesis utiliza el símbolo  $H_1$ ?

**ASISTENCIA de PH Grade** **9.3** ¿Qué símbolo utiliza para el nivel de significancia o posibilidad de cometer un error de tipo I?

**9.4** ¿Qué símbolo utiliza para la posibilidad de cometer un error de tipo II?

**ASISTENCIA de PH Grade** **9.5** ¿Qué representa  $1 - \beta$ ?

**9.6** ¿Cuál es la relación de  $\alpha$  con el error tipo I?

**9.7** ¿Cuál es la relación de  $\beta$  con el error tipo II?

**9.8** ¿Cómo se relaciona la potencia con la probabilidad de cometer un error tipo II?

**ASISTENCIA de PH Grade** **9.9** ¿Por qué es posible rechazar la hipótesis nula cuando es verdadera?

**ASISTENCIA de PH Grade** **9.10** ¿Por qué es posible no rechazar la hipótesis nula cuando es falsa?

**9.11** Para un tamaño de muestra dado, si  $\alpha$  se reduce de 0.05 a 0.01, ¿qué pasará con  $\beta$ ?

**9.12** Para  $H_0: \mu = 100$ ,  $H_1: \mu \neq 100$ , y para un tamaño de muestra de  $n$ , ¿por qué  $\beta$  es más grande si el valor real de  $\mu$  es 90, que si el valor real de  $\mu$  es 75?

### Aplicación de conceptos

**ASISTENCIA de PH Grade** **9.13** En el sistema legal estadounidense, al acusado se le considera inocente hasta que se demuestre que es culpable. Considere una hipótesis nula  $H_0$ , en la que el acusado es inocente, y una hipótesis alternativa  $H_1$ , en la que el acusado es culpable. Un jurado tiene dos posibles decisiones: encarcelar al acusado (es decir, rechazar la hipótesis nula) o exonerarlo (es decir, no rechazar la hipótesis nula). Explique el significado de los riesgos de cometer un error tipo I o un error tipo II en este ejemplo.

**9.14** Suponga que al acusado del problema 9.13 se le considera culpable hasta demostrar su inocencia, como ocurre en otros sistemas judiciales. ¿En qué difieren estas hipótesis nula y alternativa de las establecidas en el problema 9.13? ¿Cuáles son aquí los significados de los riesgos de cometer un error tipo I o un error tipo II?

**9.15** La Food and Drug Administration de Estados Unidos (FDA) es responsable de aprobar los nuevos medicamentos. Muchos grupos de consumidores creen que el proceso de aprobación es muy sencillo y que, por eso, se han aprobado muchos medicamentos que después resultan inseguros. Por otra parte, hay un buen número de cabilderos de la industria farmacéutica que pugnan por un proceso de aprobación más complaciente, de manera que a las empresas farmacéuticas se les aprueben medicinas nuevas con más facilidad y rapidez (Rochelle Sharpe, ‘FDA Tries to Find Right Balance on Drug Approvals’, *The Wall Street Journal*, 20 de abril, 1999, A24). Considere una hipótesis nula que establece que un medicamento nuevo aún sin aprobar es inseguro, y una hipótesis alternativa que establece que un medicamento nuevo aún sin aprobar es seguro.

- Explique los riesgos de cometer un error tipo I o tipo II.
- ¿Qué tipo de error tratan de evitar los grupos de consumidores? Explique su respuesta.
- ¿Qué tipo de error tratan de evitar los grupos de cabilderos? Explique su respuesta.
- ¿Cómo sería posible reducir las posibilidades de cometer los errores tipo I y tipo II?

**ASISTENCIA de PH Grade** **9.16** Como consecuencia de las quejas de alumnos y maestros en relación con sus retrasos, el decano de una gran universidad pretende ajustar los horarios de clase programados, con el fin de dejar un lapso adecuado de traslado entre clases, y está listo para emprender un estudio. Hasta ahora, el decano ha considerado que 20 minutos entre una clase y otra debe ser suficiente. Elabore las hipótesis nula  $H_0$  y alternativa  $H_1$ .

**9.17** El gerente de producción de una fábrica de telas necesita determinar si una máquina recién adquirida está produciendo cierto tipo específico de tela de acuerdo con las especificaciones de la empresa, las cuales señalan que debe tener una resistencia a la ruptura de 70 libras. Una muestra de 49 pedazos de tela revela una resistencia muestral media a la ruptura de 69.1 libras. Determine las hipótesis nula y alternativa.

**9.18** El gerente de una tienda de pinturas quiere determinar si la cantidad de pintura que contienen los envases de un galón adquiridos a un reconocido fabricante realmente promedian 1 galón. Se sabe que las especificaciones del fabricante establecen que la desviación estándar para la cantidad de pintura es de 0.02 galones. Selecciona una muestra aleatoria de 50 envases de un galón, y la media muestral resulta de 0.995 galones. Determine las hipótesis nula y alternativa.

**9.19** El gerente de control de calidad de una fábrica de focos debe determinar si la vida media de un gran lote de focos es igual al valor especificado de 375 horas. Se conoce que la desviación estándar de la población es de 100 horas. Una muestra compuesta por 64 focos indica que la vida media de la muestra es de 350 horas. Determine las hipótesis nula y alternativa.

## 9.2 PRUEBA Z DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ CONOCIDA)

Ahora que ya estudió la metodología de la prueba de hipótesis, recuerde que en el escenario “Uso de la estadística” de la página 272, la Oxford Cereal Company quiere determinar si el proceso de llenado del cereal funciona de manera adecuada (es decir, si el llenado medio a lo largo de todo el proceso de empaque continúa en los 368 gramos especificados y no se requiere acción correctiva). Para evaluar el requisito de 368 gramos, usted toma una muestra aleatoria de 25 cajas, pesa cada una, y luego evalúa la diferencia existente con el estadístico de muestra y el parámetro poblacional establecido en la hipótesis, comparando para ello el peso medio (en gramos) de la muestra con la media esperada de 368 gramos especificada por la empresa. Para este proceso de llenado, las hipótesis nula y alternativa son:

$$H_0: \mu = 368$$

$$H_1: \mu \neq 368$$

Cuando se conoce la desviación estándar,  $\sigma$ , y si la población tiene una distribución normal, usted utiliza la prueba Z. Si la población no tiene una distribución normal, todavía se puede utilizar la prueba Z si el tamaño de la muestra es lo bastante grande como para que tenga efecto el teorema del límite central (vea la sección 7.2). La ecuación (9.1) define al **estadístico de prueba Z** para determinar la diferencia que existe entre la media muestral  $\bar{X}$  y la media poblacional  $\mu$  cuando se conoce la desviación estándar  $\sigma$ .

### PRUEBA Z DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ CONOCIDA)

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (9.1)$$

En la ecuación (9.1), el numerador mide qué tan lejos (en valor absoluto) está la media muestral observada  $\bar{X}$  de la media hipotetizada  $\mu$ . El denominador es el error estándar de la media, por lo que Z representa la diferencia que existe entre  $\bar{X}$  y  $\mu$  en unidades del error estándar.

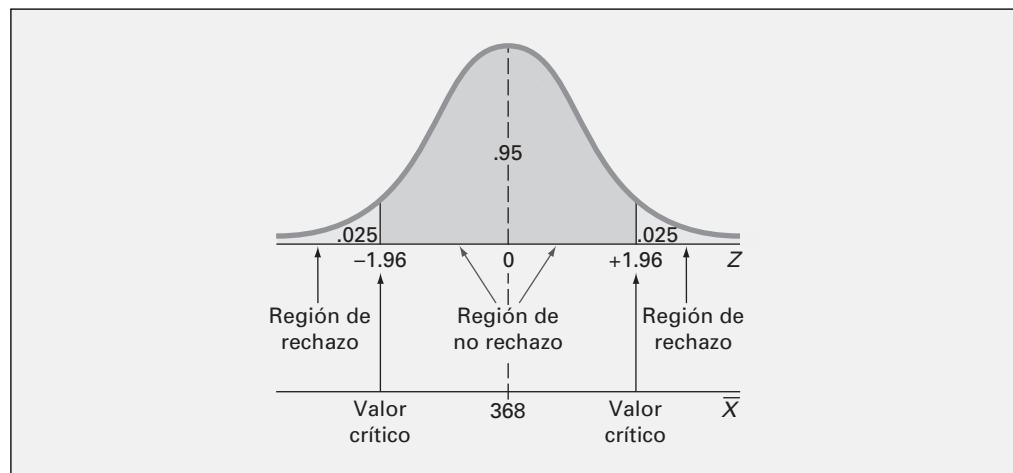
### Método del valor crítico para la prueba de hipótesis

El valor observado del estadístico de prueba Z, ecuación (9.1), se compara con los valores críticos. Estos valores críticos se expresan como valores Z estandarizados (es decir, en unidades de desviación estándar). Por ejemplo, si usted utiliza un nivel de significancia de 0.05, el tamaño de la región de rechazo es 0.05. Puesto que la región de rechazo se divide en dos colas de la distribución (esto se denomina **prueba de dos colas**), se divide el 0.05 en dos partes iguales de 0.025 cada una. Una región de rechazo de 0.025 en cada cola de la distribución normal tiene como resultado un área acumulada de 0.025 bajo el valor crítico inferior y un área acumulada de 0.975 bajo el valor crítico superior. De acuerdo con la tabla de distribución normal estandarizada acumulativa (tabla E.2), los valores críticos que dividen las regiones de rechazo y aceptación son  $-1.96$  y  $+1.96$ . La figura 9.2 ilustra que si la media es realmente de 368 gramos, como afirma  $H_0$ , entonces los valores del estadístico Z tienen una distribución normal estandarizada centrada en  $Z = 0$  (la cual corresponde a un

valor  $\bar{X}$  de 368 gramos). Los valores de  $Z$  mayores que  $+1.96$  o menores que  $-1.96$  indican que  $\bar{X}$  está muy alejada de la  $\mu = 368$  hipotetizada, por lo que sería muy poco probable que se presentara uno de estos valores si  $H_0$  fuera verdadera.

**FIGURA 9.2**

Prueba de una hipótesis sobre la media ( $\sigma$  conocida) con un nivel de significancia de 0.05.



Por lo tanto, la regla de decisión es:

Rechace  $H_0$  si  $Z > +1.96$

o si  $Z < -1.96$ ;

de lo contrario, no rechace  $H_0$ .

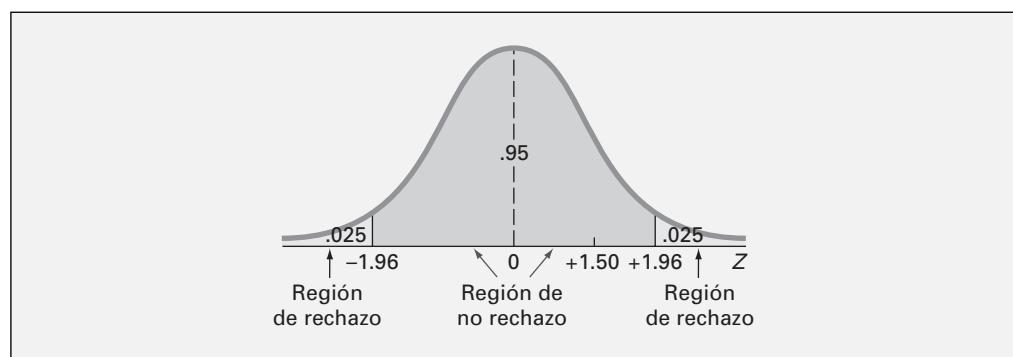
Suponga que la muestra de 25 cajas de cereal indica una media muestral de  $\bar{X} = 372.5$  gramos y que la desviación estándar de la población es de 15 gramos. Utilizando la ecuación (9.1) de la página 278:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{372.5 - 368}{\frac{15}{\sqrt{25}}} = +1.50$$

Puesto que el estadístico de prueba  $Z = +1.50$  se encuentra entre  $-1.96$  y  $+1.96$ , usted no rechaza  $H_0$  (vea la figura 9.3). Usted sigue creyendo que la cantidad media de llenado es de 368 gramos. Para tomar en cuenta la posibilidad de un error tipo II, redacta la conclusión como “no existe suficiente evidencia de que el llenado medio sea diferente a 368 gramos”.

**FIGURA 9.3**

Prueba de una hipótesis sobre la media ( $\sigma$  conocida) con un nivel de significancia de 0.05.



La exposición 9.1 muestra un resumen del método del valor crítico para la prueba de hipótesis.

### EXPOSICIÓN 9.1: MÉTODO DE SEIS PASOS DE LA PRUEBA DE HIPÓTESIS

1. Prepare la hipótesis nula,  $H_0$ , y la hipótesis alternativa  $H_1$ .
2. Seleccione el nivel de significancia,  $\alpha$ , y el tamaño de la muestra,  $n$ . El nivel de significancia se especifica de acuerdo con la importancia relativa de los riesgos de cometer errores tipo I y tipo II en el problema.
3. Determine el estadístico de prueba y la distribución muestral apropiados.
4. Determine los valores críticos que dividen las zonas de rechazo y aceptación.
5. Recopile los datos y calcule el valor del estadístico de prueba.
6. Tome la decisión estadística y establezca la conclusión administrativa. Si el estadístico de prueba queda en la región de no rechazo, usted no rechaza la hipótesis nula  $H_0$ . Si el estadístico de prueba queda en la región de rechazo, usted rechaza la hipótesis nula. La conclusión administrativa se escribe en el contexto del problema real.

### EJEMPLO 9.2

#### APLICACIÓN DEL MÉTODO DE SEIS PASOS DE PRUEBA DE HIPÓTESIS EN LA OXFORD CEREAL COMPANY

Aplique el método de seis pasos de prueba de hipótesis en la Oxford Cereal Company.

#### SOLUCIÓN

**Paso 1:** Determine las hipótesis nula y alternativa. La hipótesis nula  $H_0$  siempre se determina en términos estadísticos utilizando parámetros de la población. Al probar si el llenado medio es de 368 gramos, la hipótesis nula establece que  $\mu$  es igual a 368. La hipótesis alternativa  $H_1$ , también se determina en términos estadísticos utilizando parámetros de población. Por lo tanto, la hipótesis alternativa establece que  $\mu$  no es igual a 368 gramos.

**Paso 2:** Seleccione el nivel de significancia y el tamaño de la muestra. Usted selecciona el nivel de significancia de acuerdo con la importancia relativa de los riesgos de cometer errores tipo I y tipo II en el problema. Cuanto más pequeño es el valor de  $\alpha$ , existe menos riesgo de cometer un error tipo I. En este ejemplo, un error tipo I concluiría que la media poblacional no es de 368 gramos, cuando sí es de 368 gramos. Aquí, seleccione  $\alpha = 0.05$ . La muestra  $n = 25$ .

**Paso 3:** Seleccione el estadístico de muestra apropiado. Puesto que  $\sigma$  se conoce partir de información sobre el proceso de llenado, usted utiliza la distribución normal y el estadístico de prueba  $Z$ .

**Paso 4:** Determine la región de rechazo. Seleccione valores críticos para el estadístico de prueba apropiado, de tal manera que la región de rechazo abarque un área total de  $\alpha$  cuando  $H_0$  es cierta, y la región de aceptación abarque un área total de  $1 - \alpha$  cuando  $H_0$  es cierta. Como en el ejemplo del cereal  $\alpha = 0.05$ , los valores críticos del estadístico de prueba  $Z$  son  $-1.96$  y  $+1.96$ . La región de rechazo es  $Z < -1.96$  o  $Z > +1.96$ . La región de no rechazo es  $-1.96 < Z < +1.96$ .

**Paso 5:** Recopile los datos y calcule el valor del estadístico de prueba. En el ejemplo del cereal,  $\bar{X} = 372.5$  y el valor del estadístico de prueba es  $Z = +1.50$ .

**Paso 6:** Tome la decisión estadística y establezca la conclusión administrativa. Primero, determine si el estadístico de prueba cae dentro de la región de rechazo o no rechazo. Para el ejemplo del cereal,  $Z = +1.50$  está en la región de no rechazo, porque  $-1.96 < Z = +1.50 < +1.96$ . Puesto que el estadístico de prueba queda en la zona de no rechazo, la decisión estadística pertinente es no rechazar la hipótesis nula  $H_0$ . La conclusión administrativa radica en que no existe evidencia suficiente para demostrar que, con un nivel de significancia de 0.05, el llenado medio es distinto de 368 gramos. No se necesita acción correctiva.

**EJEMPLO 9.3****RECHAZAR UNA HIPÓTESIS NULA**

Usted es gerente de un restaurante de comida rápida. Quiere determinar si el tiempo de espera al pedir una orden se ha modificado durante el último mes con respecto a su valor de media poblacional previo de 4.5 minutos. A partir de la experiencia anterior, supone que la desviación estándar de la población es de 1.2 minutos. Selecciona una muestra de 25 órdenes durante un periodo de una hora. La media muestral es de 5.1 minutos. Utilice el método de seis pasos de la exposición 9.1 de la página 280, para determinar si existe evidencia de que, con un nivel de significancia de 0.05, el tiempo de espera medio para servir una orden se ha modificado durante el último mes con respecto a su valor de media poblacional previo de 4.5 minutos.

**SOLUCIÓN**

**Paso 1:** La hipótesis nula dice que la media poblacional no ha cambiado con respecto a su valor previo de 4.5 minutos.

$$H_0: \mu = 4.5$$

La hipótesis alternativa es opuesta a la hipótesis nula. Puesto que la hipótesis nula dice que la media poblacional es de 4.5 minutos, la hipótesis alternativa dice que la media poblacional no es de 4.5 minutos.

$$H_1: \mu \neq 4.5$$

**Paso 2:** Usted ha seleccionado una muestra de  $n = 25$ . El nivel de significancia es 0.05 (es decir,  $\alpha = 0.05$ ).

**Paso 3:** Puesto que conoce  $\sigma$ , usted utiliza la distribución normal y el estadístico de prueba Z.

**Paso 4:** Como  $\alpha = 0.05$ , los valores críticos del estadístico de prueba Z son  $-1.96$  y  $+1.96$ . La región de rechazo es  $Z < -1.96$  o  $Z > +1.96$ . La zona de no rechazo es  $-1.96 < Z < +1.96$ .

**Paso 5:** Usted recopila los datos y calcula  $\bar{X} = 5.1$ . Utilizando la ecuación (9.1) de la página 278, calcule el estadístico de prueba.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.1 - 4.5}{\frac{1.2}{\sqrt{25}}} = 2.50$$

**Paso 6:** Puesto que  $Z = 2.50 > 1.96$ , usted rechaza la hipótesis nula. Concluye que existe evidencia de que el tiempo de espera al pedir una orden se ha modificado con respecto a su valor de media poblacional previo de 4.5 minutos. El tiempo de espera medio de los clientes ahora es mayor que el mes pasado.

**Método del valor-p para la prueba de hipótesis**

La mayoría de los programas de cómputo moderno, incluyendo Excel, Minitab y SPSS, calculan el valor-*p* al realizar una prueba de hipótesis.

El **valor-*p*** es la probabilidad de obtener un estadístico de prueba igual o más extremo que el resultado de la muestra, dado que la hipótesis nula  $H_0$  es cierta.

El valor-*p*, que a menudo se denomina *nivel de significancia observado*, es el nivel más pequeño en el que se puede rechazar  $H_0$ .

Las reglas de decisión para rechazar  $H_0$  en el método del valor-*p* son:

- Si el valor-*p* es mayor o igual que  $\alpha$ , no rechace la hipótesis nula.
- Si el valor-*p* es menor que  $\alpha$ , rechace la hipótesis nula.

Mucha gente confunde estas reglas, creyendo equivocadamente que un valor-*p* grande es argumento para el rechazo. Usted evitará esta confusión si memoriza la siguiente regla.

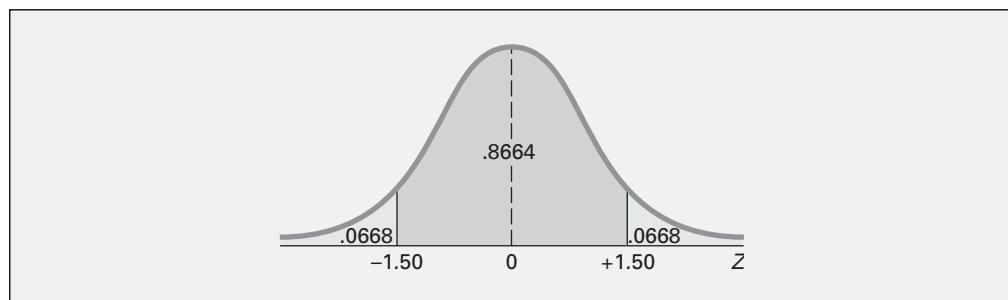
Si el valor-*p* es bajo, entonces  $H_0$  debe irse.

Para comprender mejor el método del valor-*p*, consideraremos de nuevo el escenario de la Oxford Cereal Company. Usted probó si el llenado medio era igual o no a 368 gramos. El estadístico de prueba tuvo como resultado un valor *Z* de +1.50 y usted no rechazó la hipótesis nula porque +1.50 era menor que el valor crítico superior de +1.96 y mayor que el valor crítico inferior de -1.96.

Para utilizar el método del valor-*p* en la *prueba de dos colas*, encuentre la probabilidad de obtener un estadístico de prueba *Z* igual o *más extremo que* 1.50 unidades de desviación estándar con respecto al centro de una distribución normal estandarizada. En otras palabras, es necesario calcular la probabilidad de un valor *Z* mayor que +1.50 junto con la probabilidad de un valor *Z* menor que -1.50. La tabla E.2 muestra que la probabilidad de un valor *Z* inferior a -1.50 es de 0.0668. La probabilidad de un valor inferior a +1.50 es 0.9332 y la probabilidad de un valor por arriba de +1.50 es  $1 - 0.9332 = 0.0668$ . Por lo tanto, el valor-*p* para esta prueba de dos colas es  $0.0668 + 0.0668 = 0.1336$  (vea la figura 9.4). De esta forma, la probabilidad de un resultado igual o más extremo que el observado es 0.1336. Como 0.1336 es mayor que  $\alpha = 0.05$ , no rechace la hipótesis nula.

**FIGURA 9.4**

Cálculo de un valor-*p* para una prueba de dos colas.



En este ejemplo, la media muestral observada es de 372.5 gramos, 4.5 gramos arriba del valor establecido en la hipótesis y el valor-*p* es 0.1336. De esta forma, si la media poblacional es de 368 gramos, hay un 13.36% de posibilidades de que la media muestral se encuentre a más de 4.5 gramos de 368 (es decir, mayor o igual que 372.5 gramos, o menor o igual que 363.5 gramos). Por lo tanto, aunque 372.5 se encuentra por arriba del valor de 368 hipotetizado, cuando la media poblacional es de 368, no es altamente probable un resultado tan o más extremo que 372.5.

A menos que usted esté manejando un estadístico de prueba que siga una distribución normal, el cálculo del valor-*p* resultará muy complicado. Sin embargo, algunos programas de cómputo como Excel, Minitab o SPSS exponen en forma rutinaria el valor-*p* como parte del resultado de los procedimientos de prueba de hipótesis. La figura 9.5 muestra una hoja de trabajo de Excel correspondiente al ejemplo de llenado de cereal analizado en esta sección.

**FIGURA 9.5**

Hoja de la prueba *Z* en Excel para el ejemplo de llenado de cereal.

A	B
1 Cereal-Filling Process Hypothesis Test	
2	
3 Data	
4 Null Hypothesis $\mu =$	368
5 Level of Significance $\alpha =$	0.05
6 Population Standard Deviation $\sigma =$	15
7 Sample Size $n =$	25
8 Sample Mean $\bar{x} =$	372.5
9	
10 Intermediate Calculations	
11 Standard Error of the Mean $\sigma_{\bar{x}} =$	3
12 Z Test Statistic $Z =$	1.5
13	
14 Two-Tail Test	
15 Lower Critical Value $Z_{\text{lower}} =$	-1.9600
16 Upper Critical Value $Z_{\text{upper}} =$	1.9600
17 <i>p</i> Value $p =$	0.1336
18 Do not reject the null hypothesis	
	=B6/RAIZ(B7)
	=(B8 - B4)/B11
	=DISTR.NORM.ESTAND.INV(1B5/2)
	=DISTR.NORM.ESTAND.INV(1 - B5/2)
	=2 * (1 - DISTR.NORM.ESTAND.(ABS(B12)))

La exposición 9.2 muestra un resumen del método del valor- $p$  para la prueba de hipótesis.

**EXPOSICIÓN 9.2: MÉTODO DE CINCO PASOS PARA LA PRUEBA DE HIPÓTESIS UTILIZANDO EL MÉTODO DEL VALOR- $P$**

1. Prepare la hipótesis nula,  $H_0$ , y la hipótesis alternativa  $H_1$ .
2. Seleccione el nivel de significancia,  $\alpha$ , y el tamaño de la muestra,  $n$ . El nivel de significancia se especifica de acuerdo con la importancia relativa de los riesgos de cometer errores tipo I y tipo II en el problema.
3. Determine el estadístico de prueba y la distribución muestral apropiados.
4. Recopile los datos, encuentre el valor del estadístico de prueba y calcule el valor- $p$ .
5. Tome la decisión estadística y establezca la conclusión administrativa. Si el valor- $p$  es mayor o igual que  $\alpha$ , no rechace la hipótesis nula  $H_0$ . Si el valor- $p$  es menor que  $\alpha$ , rechace la hipótesis nula. Recuerde la frase: si el valor- $p$  es bajo, entonces  $H_0$  debe irse. La conclusión administrativa se escribe en el contexto del problema real.

**EJEMPLO 9.4**

**RECHAZAR UNA HIPÓTESIS NULA UTILIZANDO EL MÉTODO DEL VALOR- $P$**

Usted es gerente de un restaurante de comida rápida. Quiere determinar si el tiempo de espera al pedir una orden se ha modificado durante el último mes con respecto a su valor de media poblacional previo de 4.5 minutos. A partir de la experiencia anterior, supone que la desviación estándar de la población es de 1.2 minutos. Selecciona una muestra de 25 órdenes durante un periodo de una hora. La media muestral es de 5.1 minutos. Utilice el método de cinco pasos de la exposición 9.2 ya mencionada, para determinar si existe evidencia de que el tiempo de espera medio para servir una orden se ha modificado durante el último mes con respecto a su valor medio poblacional previo de 4.5 minutos.

**SOLUCIÓN**

**Paso 1:** La hipótesis nula dice que la media poblacional no ha cambiado con respecto a su valor previo de 4.5 minutos.

$$H_0: \mu = 4.5$$

La hipótesis alternativa es opuesta a la hipótesis nula. Puesto que la hipótesis nula dice que la media poblacional es de 4.5 minutos, la hipótesis alternativa dice que la media poblacional no es de 4.5 minutos.

$$H_1: \mu \neq 4.5$$

**Paso 2:** Usted seleccionó un tamaño de la muestra de  $n = 25$  y elige un nivel de significancia de 0.05 (es decir,  $\alpha = 0.05$ ).

**Paso 3:** Seleccione el estadístico de prueba apropiado. Puesto que conoce  $\sigma$ , usted utiliza la distribución normal y el estadístico de prueba  $Z$ .

**Paso 4:** Recopile los datos y calcule  $\bar{X} = 5.1$ . Mediante la ecuación (9.1) de la página 278, calcule el estadístico de prueba.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5.1 - 4.5}{\frac{1.2}{\sqrt{25}}} = 2.50$$

Para encontrar la probabilidad de obtener un estadístico de prueba  $Z$  igual o *más extremo que* 2.50 unidades de desviación estándar con respecto al centro de una distribución normal estandarizada, usted calcula la probabilidad de un valor  $Z$  mayor que 2.50 junto con la

probabilidad de un valor  $Z$  menor que  $-2.50$ . A partir de la tabla E.2, la probabilidad de un valor  $Z$  por debajo de  $-2.50$  es  $0.0062$ . La probabilidad de un valor por debajo de  $+2.50$  es  $0.9938$ . Por lo tanto, la probabilidad de un valor por encima de  $+2.50$  es  $1 - 0.9938 = 0.0062$ . Así, el valor- $p$  para esta prueba de dos colas es  $0.0062 + 0.0062 = 0.0124$ .

**Paso 5:** Puesto que el valor- $p = 0.0124 < \alpha = 0.05$ , usted rechaza la hipótesis nula. Concluye que existe evidencia de que el tiempo de espera al pedir una orden se ha modificado con respecto a su valor de media poblacional previo de  $4.5$  minutos. El tiempo de espera medio de los clientes ahora es mayor que el mes pasado.

## Conexión entre la estimación del intervalo de confianza y la prueba de hipótesis

En este capítulo y en el anterior se analizan los dos componentes principales de la inferencia estadística: la estimación del intervalo de confianza y la prueba de hipótesis. Aunque ambas se basan en el mismo conjunto de conceptos, se utilizan con distintos propósitos. En el capítulo 8 se utilizaron intervalos de confianza para estimar parámetros. En este capítulo, la prueba de hipótesis se utiliza para tomar decisiones respecto a valores específicos de los parámetros poblacionales. Las prueba de hipótesis se utilizan al tratar de comprobar que un parámetro es menor que, mayor que, o diferente que un valor especificado. Sin embargo, la interpretación adecuada de un intervalo de confianza también indica si un parámetro es menor que, mayor que, o diferente que un valor especificado.

Por ejemplo, en esta sección usted probó si la media poblacional era diferente a  $368$  gramos, mediante el uso de la ecuación (9.1) de la página 278:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

En lugar de probar la hipótesis nula de que  $\mu = 368$  gramos, llegará a la misma conclusión si conforma un intervalo de confianza estimado de  $\mu$ . Si el valor establecido en la hipótesis de  $\mu = 368$  queda dentro del intervalo, no rechace la hipótesis nula porque  $368$  no se considera un valor inusual. Por otra parte, si el valor hipotetizado no queda dentro del intervalo, rechace la hipótesis nula porque “ $368$  gramos” entonces se considera un valor inusual. Al emplear la ecuación (8.1) de la página 241 y los datos siguientes:

$$n = 25, \bar{X} = 372.5 \text{ gramos}, \sigma = 15 \text{ gramos}$$

Para un nivel de confianza del  $95\%$  (correspondiente a un nivel de significancia de  $0.05$  —es decir,  $\alpha = 0.05$ ),

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

$$372.5 \pm (1.96) \frac{15}{\sqrt{25}}$$

$$372.5 \pm 5.88$$

de manera que

$$366.62 \leq \mu \leq 378.38$$

Puesto que el intervalo incluye al valor de  $368$  gramos establecidos en la hipótesis, no rechace la hipótesis nula. No existe evidencia suficiente de que la media de la cantidad de llenado durante todo el proceso no sea de  $368$  gramos. Usted obtuvo la misma decisión mediante el uso de la prueba de hipótesis.

## PROBLEMAS PARA LA SECCIÓN 9.2

### Aprendizaje básico

**ASISTENCIA de PH Grade** **9.20** Si usted utiliza un nivel de significancia de 0.05 en una prueba de hipótesis (de dos colas), ¿qué decidiría si el valor del estadístico de prueba  $Z$  calculado fuera  $+2.21$ ?

**ASISTENCIA de PH Grade** **9.21** Si usted utiliza un nivel de significancia de 0.10 en una prueba de hipótesis (de dos colas), ¿cuál sería su regla de decisión para rechazar una hipótesis nula donde la media poblacional es 500 si utiliza la prueba  $Z$ ?

**ASISTENCIA de PH Grade** **9.22** Si usted utiliza un nivel de significancia de 0.10 en una prueba de hipótesis (de dos colas), ¿cuál es su regla de decisión para rechazar una  $H_0: \mu = 12.5$ , si utiliza la prueba  $Z$ ?

**ASISTENCIA de PH Grade** **9.23** ¿Cuál sería su decisión al problema 9.22, si el valor del estadístico de prueba  $Z$  calculado es  $-2.61$ ?

**ASISTENCIA de PH Grade** **9.24** Suponga que en una prueba de hipótesis de dos colas calcula que el valor del estadístico de prueba  $Z$  es  $+2.00$ . ¿Cuál es el valor- $p$ ?

**ASISTENCIA de PH Grade** **9.25** En el problema 9.24, ¿cuál sería su decisión estadística si prueba la hipótesis nula con un nivel de significancia de 0.10?

**ASISTENCIA de PH Grade** **9.26** Suponga que en una prueba de hipótesis de dos colas calcula que el valor del estadístico de prueba  $Z$  es  $-1.38$ . ¿Cuál es el valor- $p$ ?

**ASISTENCIA de PH Grade** **9.27** En el problema 9.26, ¿cuál sería su decisión estadística si prueba la hipótesis nula con un nivel de significancia de 0.01?

### Aplicación de conceptos

**AUTO Examen** **9.28** El gerente de producción de una fábrica de telas necesita determinar si una máquina recién adquirida está produciendo cierto tipo específico de tela de acuerdo con las especificaciones de la empresa, las cuales señalan que debe tener una resistencia a la ruptura de 70 libras y una desviación estándar de 3.5 libras. Una muestra de 49 pedazos de tela revela una resistencia muestral media a la ruptura de 69.1 libras.

- Existe evidencia de que la máquina no está cumpliendo con las especificaciones del fabricante en cuanto a la resistencia media a la ruptura? (Utilice un nivel de significancia de 0.05.)
- Calcule el valor- $p$  e interprete su significado.
- ¿Cuál sería su respuesta al inciso a), si la desviación estándar es de 1.75 libras?
- ¿Cuál sería su respuesta al inciso a), si la media muestral es de 69 libras y la desviación estándar es de 3.5 libras?

**ASISTENCIA de PH Grade** **9.29** El gerente de una tienda de pinturas quiere determinar si la cantidad de pintura que contienen los envases de un galón adquiridos a un reconocido fabricante es realmente de un galón. Usted sabe que las especifi-

caciones del fabricante establecen que la desviación estándar de la cantidad de pintura es de 0.02 galones. Selecciona una muestra aleatoria de 50 envases, y la cantidad media de pintura por envase de 1 galón resulta de 0.995 galones.

- Existe evidencia de que la cantidad media es diferente de 1.0 galones (utilice  $\alpha = 0.01$ )?
- Calcule el valor- $p$  e interprete su significado.
- Construya un intervalo de confianza estimado del 99% para la cantidad media poblacional de pintura.
- Compare los resultados de a) y c). ¿A qué conclusiones llega?

**9.30** El gerente de control de calidad de una fábrica de focos debe determinar si la vida media de un gran lote de focos es igual al valor especificado de 375 horas. La desviación estándar de la población es 100 horas. Una muestra compuesta por 64 focos indica una vida media muestral de 350 horas.

- Con un nivel de significancia de 0.05, ¿existe evidencia de que la vida media es distinta de 375 horas?
- Calcule el valor- $p$  e interprete su significado.
- Elabore un intervalo de confianza estimado del 95% de la vida media poblacional de los focos.
- Compare los resultados de a) y c). ¿A qué conclusiones llega?

**9.31** La división de inspectores del Departamento de Pesos y Medidas del condado de Lee está interesada en determinar si en las botellas de 2 litros procesadas en la planta embotelladora local, perteneciente a una reconocida y gran empresa, se ha colocado la cantidad apropiada de bebida gaseosa. La embotelladora informó a la división de inspectores que la desviación estándar de las botellas de 2 litros es de 0.05. Una muestra aleatoria conformada por 100 botellas de dos litros, tomada de la planta embotelladora, señala una media muestral de 1.99 litros.

- Con un nivel de significancia de 0.05, ¿existe evidencia de que la cantidad media en las botellas es distinta de 2 litros?
- Calcule el valor- $p$  e interprete su significado.
- Elabore un intervalo de confianza estimado del 95% de la cantidad media poblacional en las botellas.
- Compare los resultados de los incisos a) y c). ¿A qué conclusiones llega?

**9.32** Un fabricante de aderezos para ensalada utiliza máquinas para suministrar ingredientes líquidos a las botellas que pasan por la línea de llenado. La máquina que suministra los aderezos está funcionando de manera apropiada cuando la cantidad media abastecida es de 8 onzas. La desviación estándar poblacional de la cantidad abastecida es de 0.15 onzas. Periódicamente se selecciona una muestra de 50 botellas y, si se encuentran evidencias de que la cantidad media suministrada es distinta de 8 onzas, se detiene la línea de llenado. Suponga que la cantidad media abastecida a una muestra en particular de 50 botellas es 7.983 onzas.

- Existe evidencia de que la cantidad media poblacional es diferente de 8 onzas? (Utilice un nivel de significancia de 0.05.)
- Calcule el valor- $p$  e interprete su significado.

- c. ¿Cuál sería su respuesta al inciso a), si la desviación estándar fuera de 0.05 onzas?
- d. ¿Cuál sería su respuesta al inciso a), si la media muestral fuera de 7.952 onzas y la desviación estándar de 0.15 onzas?

**9.33** Los cajeros automáticos deben contar con efectivo suficiente para satisfacer los retiros de los clientes durante todo el fin de semana. Pero si se deja en ellos demasiado efectivo innecesariamente, el banco se priva de la oportunidad de invertir ese dinero y ganar intereses. Suponga que en una sucursal específica la cantidad media poblacional de dinero retirado del cajero automático por transacción durante el fin de semana es

de 160 dólares, con una desviación estándar poblacional de 30 dólares.

- a. Si una muestra aleatoria de 36 transacciones indica que la media muestral de la cantidad retirada es de 172 dólares, ¿existen evidencias para creer que la media poblacional de la cantidad retirada no es mayor que 160 dólares? (Utilice un nivel de significancia de 0.05.)
- b. Calcule el valor- $p$  e interprete su significado.
- c. ¿Cuál sería su respuesta al inciso b) si utiliza un nivel de significancia de 0.01?
- d. ¿Cuál sería su respuesta al inciso b) si la desviación estándar es de 24 dólares (utilice  $\alpha = 0.05$ )?

## 9.3 PRUEBAS DE UNA COLA

Hasta ahora, la metodología de prueba de hipótesis se ha utilizado para examinar la interrogante sobre si la cantidad media poblacional de cereal envasado es o no de 368 gramos. La hipótesis alternativa ( $H_1: \mu \neq 368$ ) abarca dos posibilidades: que la media sea menor que 368 gramos, o que la media sea mayor que 368 gramos. Por ello, la región de rechazo se divide en las dos colas de la distribución muestral de la media.

Sin embargo, en muchos casos la hipótesis alternativa se orienta en una *dirección en particular*. En la siguiente situación se presenta uno de estos casos. Una empresa que fabrica queso procesado está interesada en determinar si algunos de sus proveedores de leche le están añadiendo agua para aumentar la cantidad suministrada a la operación de procesamiento. Usted sabe que el exceso de agua reduce el punto de congelación de la leche. El punto de congelación de la leche natural tiene una distribución normal, con una media de  $-0.545^\circ$  Celsius (C). La desviación estándar de la temperatura de congelación de leche natural se sabe de  $0.008^\circ$  C. Puesto que la empresa productora de queso sólo está interesada en determinar si el punto de congelación de la leche es menor que el que cabría esperar de la leche natural, toda la región de rechazo se localiza en la cola inferior de la distribución.

### Método del valor crítico

Supongamos que quiere determinar si el punto medio de congelación de la leche es menor que  $-0.545^\circ$ . Para realizar esta prueba de hipótesis de una cola, utiliza el método de seis pasos de la exposición 9.1 de la página 280.

**Paso 1:**  $H_0: \mu \geq -0.545^\circ$   
 $H_1: \mu < -0.545^\circ$

La hipótesis alternativa contiene la afirmación que usted está tratando de comprobar. Si rechaza la hipótesis nula, existe una demostración estadística de que el punto medio de congelamiento es menor que el punto de congelación natural de  $-0.545^\circ$ . Si la conclusión de la prueba es “no rechace  $H_0$ ”, entonces no existe evidencia suficiente para comprobar que el punto medio de congelación está por debajo del punto de congelación natural de  $-0.545^\circ$ .

**Paso 2:** Usted seleccionó un tamaño de la muestra de  $n = 25$ . Decide utilizar  $\alpha = 0.05$ .

**Paso 3:** Puesto que conoce  $\sigma$ , usted utiliza la distribución normal y el estadístico de prueba  $Z$ .

**Paso 4:** La región de rechazo se encuentra totalmente comprendida dentro de la cola inferior de la distribución muestral de la media, ya que usted quiere rechazar  $H_0$  sólo cuando la media muestral sea notablemente inferior a  $-0.545^\circ$ . Cuando toda la región de rechazo se encuentra dentro de una cola de la distribución muestral del estadístico de prueba, la prueba se denomina **prueba de una cola o prueba direccional**. Cuando la hipótesis alternativa incluye el signo *menor que*, el valor crítico de  $Z$  debe ser menor que 0. Como indican la tabla 9.2 y la figura 9.6, el valor crítico del estadístico de prueba  $Z$  es  $-1.645$ , la media de  $-1.64$

y  $-1.65$ , puesto que toda la región de rechazo está en la cola inferior de la distribución normal estandarizada y contiene un área de  $0.05$ .

La regla de decisión es:

Rechace  $H_0$  si  $Z < -1.645$ ;

de otra manera, no rechace  $H_0$ .

**TABLA 9.2**

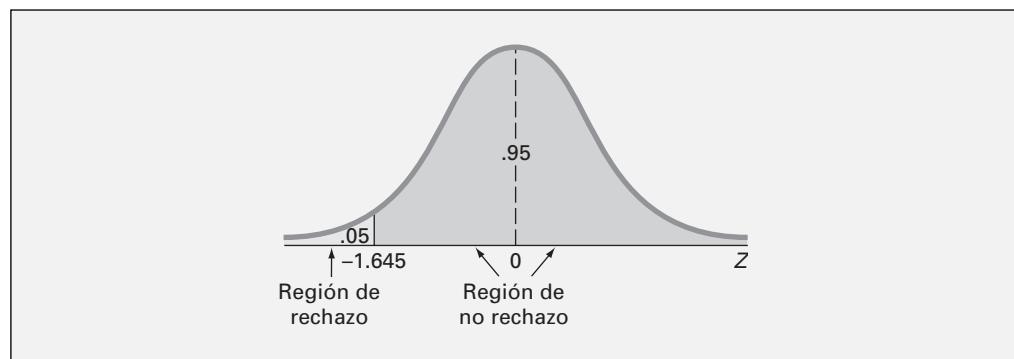
Cálculo del valor crítico del estadístico de prueba  $Z$  a partir de la distribución normal estandarizada, para una prueba de una cola con  $\alpha = 0.05$ .

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
o	o	o	o	o	o	o	o	o	o	o
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455

Fuente: Sintetizado de la tabla E.2.

**FIGURA 9.6**

Prueba de hipótesis de una cola para una media ( $\sigma$  conocida) con un nivel de significancia de  $0.05$ .



**Paso 5:** Usted selecciona una muestra de 25 envases de leche y calcula que el punto medio de congelación muestral es de  $-0.550^\circ$ . Utilizando  $n = 25$ ,  $\bar{X} = -0.550^\circ$ ,  $\sigma = 0.008^\circ$ , y la ecuación (9.1) de la página 278,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{-0.550 - (-0.545)}{\frac{0.008}{\sqrt{25}}} = -3.125$$

**Paso 6:** Puesto que  $Z = -3.125 < -1.645$ , usted rechaza la hipótesis nula (vea la figura 9.6) y concluye que el punto medio de congelación de la leche suministrada está por debajo de  $-0.545^\circ$ . La empresa debe emprender una investigación del proveedor de leche, porque el punto medio de congelación se encuentra muy por debajo de lo que cabría esperar como posibilidad.

### Método del valor-*p*

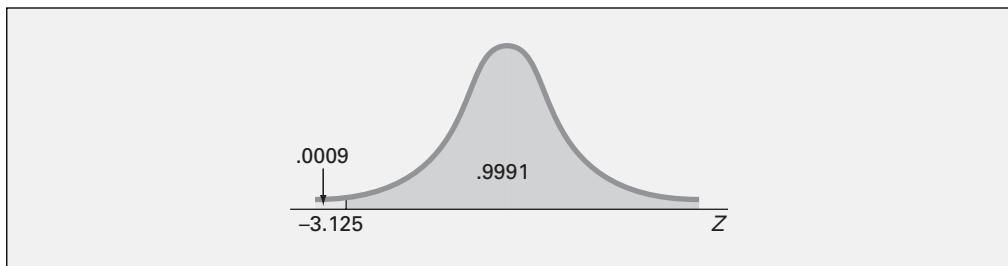
Utilice los cinco pasos numerados en la exposición 9.2 de la página 283 para ilustrar el ejemplo anterior utilizando el método del valor-*p*.

**Paso 1-3:** Estos pasos se realizan exactamente igual que en el método del valor crítico.

**Paso 4:**  $Z = -3.125$  (vea el paso 5 del método del valor crítico). Puesto que la hipótesis alternativa señala una región de rechazo totalmente dentro de la cola *inferior* de la distribución muestral del estadístico de prueba  $Z$ , para calcular el valor-*p* usted necesita encontrar la probabilidad de que el valor  $Z$  se encuentre *por debajo* del estadístico de prueba de  $-3.125$ . A partir de la tabla E.2, la probabilidad de que el valor  $Z$  sea inferior a  $-3.125$  es  $0.0009$  (vea las figuras 9.7 y 9.8).

**FIGURA 9.7**

Determinación del valor-*p* en una prueba de una cola.

**FIGURA 9.8**

Resultado de la prueba Z en Excel para el ejemplo de la producción de leche.

A	B
1 Milk Production Hypothesis	
2	
3 Data	
4 Null Hypothesis $\mu =$	0.545
5 Level of Significance	0.05
6 Population Standard Deviation	0.008
7 Sample Size	25
8 Sample Mean	0.55
9	
10 Intermediate Calculations	
11 Standard Error of the Mean	0.0016
12 Z Test Statistic	-3.125
13	
14 Lower-Tail Test	
15 Lower Critical Value	-1.6449
16 p Value	0.0009
17 Reject the null hypothesis	

$=B6/RAIZ(B7)$   
 $=(B8 - B4)/B11$   
 $=DISTR.NORM.ESTAND.INV(B5)$   
 $=DISTR.NORM.ESTAND.(B12)$

**Paso 5:** El valor-*p* de 0.0009 es menor que  $\alpha = 0.05$ . Usted rechaza  $H_0$ . Concluye que el punto medio de congelación de la leche suministrada está por debajo de  $-0.545^\circ$ . La empresa debe emprender una investigación del proveedor de leche, porque el punto medio de congelación se encuentra muy por debajo de lo que cabría esperar como posibilidad.

### EJEMPLO 9.5

### PRUEBA DE UNA COLA PARA LA MEDIA

Una empresa fabricante de chocolate en tablillas está especialmente preocupada porque el peso medio de su barra de chocolate no supere las 6.03 onzas. A partir de una experiencia anterior, usted supone que la desviación estándar es de 0.02 onzas. Selecciona una muestra aleatoria de 50 barras y la media muestral resulta de 6.034 onzas. Utilizando un nivel de significancia de  $\alpha = 0.01$ , ¿existen evidencias de que el peso medio de la población de barras de chocolate es mayor que 6.03 onzas?

**SOLUCIÓN** Se utiliza el método del valor crítico:

**Paso 1:**  $H_0: \mu \leq 6.03$   
 $H_1: \mu > 6.03$

**Paso 2:** Usted seleccionó un tamaño de la muestra de  $n = 50$ . Decide utilizar  $\alpha = 0.01$ .

**Paso 3:** Puesto que conoce  $\sigma$ , usted utiliza la distribución normal y el estadístico de prueba *Z*.

**Paso 4:** La región de rechazo se encuentra totalmente comprendida dentro de la cola superior de la distribución muestral de la media, ya que se rechazará  $H_0$  sólo cuando la media muestral sea notablemente superior a 6.03 onzas. Puesto que toda la región de rechazo está en la cola superior de la distribución normal estandarizada y contiene un área de 0.01, el valor crítico del estadístico de prueba *Z* es 2.33.

La regla de decisión es:

Rechace  $H_0$  si  $Z > 2.33$ ;  
de cualquier otra manera, no rechace  $H_0$ .

**Paso 5:** Seleccionó una muestra de 50 tablillas de chocolate y el peso medio muestral resultó de 6.034 onzas. Utilizando  $n = 50$ ,  $\bar{X} = 6.034$ ,  $\sigma = 0.02$ , y la ecuación (9.1) de la página 278:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{6.034 - 6.03}{\frac{0.02}{\sqrt{50}}} = 1.414$$

**Paso 6:** Como  $Z = 1.414 < 2.33$ , no se rechaza la hipótesis nula.

No existe evidencia suficiente para concluir que el peso medio poblacional está por encima de 6.03 onzas.

Para realizar pruebas de hipótesis de una cola, hay que formular  $H_0$  y  $H_1$  de manera apropiada. A continuación se encuentra un resumen de las hipótesis nula y alternativa para pruebas de una cola.

1. La hipótesis nula  $H_0$  representa al *status quo* o creencia actual en una situación.
2. La hipótesis alternativa  $H_1$  es lo opuesto a la hipótesis nula y representa una afirmación de investigación o inferencia específica que quisiera demostrar.
3. Si usted rechaza la hipótesis nula, tiene una prueba estadística de que la hipótesis alternativa es correcta.
4. Si usted no rechaza la hipótesis nula, entonces no ha podido demostrar la hipótesis alternativa. Sin embargo, el no poder demostrar la hipótesis alternativa no significa que se haya demostrado la hipótesis nula.
5. La hipótesis nula ( $H_0$ ) siempre se refiere a un valor específico del *parámetro poblacional* (como  $\mu$ ), no a un *estadístico de muestra* (como  $\bar{X}$ ).
6. El enunciado que describe la hipótesis nula *siempre* contiene un signo de igual en relación con el valor específico del parámetro (por ejemplo,  $H_0: \mu \geq -0.545^\circ\text{C}$ ).
7. El enunciado que describe la hipótesis alternativa  *nunca* contiene un signo de igual relacionado con el valor específico del parámetro (por ejemplo,  $H_1: \mu < -0.545^\circ\text{C}$ ).

## PROBLEMAS PARA LA SECCIÓN 9.3

### Aprendizaje básico

ASISTENCIA  
de PH Grade

**9.34** ¿Cuál es el valor crítico en la *cola superior* del estadístico de prueba  $Z$ , con un nivel de significancia de 0.01?

ASISTENCIA  
de PH Grade

**9.35** En el problema 9.34, ¿cuál sería su decisión estadística si el valor calculado del estadístico de prueba  $Z$  fuera  $+2.39$ ?

ASISTENCIA  
de PH Grade

**9.36** ¿Cuál es el valor crítico en la *cola inferior* del estadístico de prueba  $Z$ , con un nivel de significancia de 0.01?

ASISTENCIA  
de PH Grade

**9.37** En el problema 9.36, ¿cuál sería su decisión estadística si el valor calculado del estadístico de prueba  $Z$  fuera  $-1.15$ ?

ASISTENCIA  
de PH Grade

**9.38** Suponga que en una prueba de hipótesis con una cola en la que se rechaza  $H_0$  sólo en la *cola superior*, se calculó que el valor del estadístico de prueba  $Z$  es  $+2.00$ . ¿Cuál es el valor- $p$ ?

ASISTENCIA  
de PH Grade

**9.39** En el problema 9.38, ¿cuál es su decisión estadística si probó la hipótesis nula con un nivel de significancia de 0.05?

ASISTENCIA  
de PH Grade

**9.40** Suponga que en una prueba de hipótesis con una cola en la que se rechaza  $H_0$  sólo en la *cola inferior*, se calculó que el valor del estadístico de prueba  $Z$  es  $-1.38$ . ¿Cuál es el valor- $p$ ?

ASISTENCIA  
de PH Grade

**9.41** En el problema 9.40, ¿cuál es su decisión estadística si probó la hipótesis nula con un nivel de significancia de 0.01?

**9.42** En una prueba de hipótesis con una cola en la que se rechaza  $H_0$  sólo en la *cola inferior*, se calculó que el valor del estadístico de prueba  $Z$  es  $+1.38$ , ¿cuál es el valor- $p$ ?

**9.43** En el problema 9.42, ¿cuál sería la decisión estadística si se probó la hipótesis nula con un nivel de significancia de 0.01?

### Aplicación de conceptos

ASISTENCIA  
de PH Grade

**9.44** La empresa Glen Valley Steel Company fabrica barras de acero. Si el proceso de producción funciona de forma adecuada, las barras de acero que se fabrican tienen una longitud media de *por lo menos* 2.8 pies, con una desviación estándar de 0.20 (como lo determinan las especificaciones de ingeniería del equipo de producción). Las barras de acero más largas se pueden utilizar o modificar, pero

las barras más cortas se tienen que desechar. Usted selecciona una muestra de 25 barras y la longitud media resulta de 2.73 pies. ¿Es necesario ajustar el equipo de producción?

- Si quiere probar la hipótesis nula con un nivel de significancia de 0.05, ¿qué decisión tomaría utilizando el método del valor crítico para probar la hipótesis?
- Si quiere probar la hipótesis nula con un nivel de significancia de 0.05, ¿qué decisión tomaría utilizando el método del valor-*p* para probar la hipótesis?
- Interprete el significado del valor-*p* en este problema.
- Compare sus conclusiones de los incisos *a*) y *b*).

**9.45** Usted es gerente de un restaurante que entrega pizzas a los dormitorios de una universidad. Acaba de modificar su proceso de entrega con la finalidad de reducir el tiempo medio transcurrido entre el pedido y la entrega, que actualmente es de 25 minutos. A partir de su experiencia anterior, supone que la desviación estándar de la población es de 6 minutos. Una muestra de 36 órdenes en las que utilizó un nuevo proceso de entrega genera una media muestral de 22.4 minutos.

- Utilizando los seis pasos del método del valor crítico, con un nivel de significancia de 0.05, ¿existe evidencia de que se ha reducido el tiempo de entrega medio, por debajo del valor previo de la media poblacional de 25 minutos?
- Utilice los cinco pasos del método del valor-*p*, con un nivel de significancia de 0.05.
- Interprete el significado del valor-*p* en el inciso *b*).
- Compare sus conclusiones de los incisos *a*) y *b*).

**9.46** En Estados Unidos, los niños son responsables por ventas que ascienden a 36 mil millones de dólares al año. Cuando se considera su influencia directa en la elección de productos, desde estéreos hasta vacaciones, el gasto económico total en el que influyen los niños en Estados Unidos es de 290 mil millones de dólares. Se estima que a los 10 años, un niño realiza un promedio de más de cinco salidas a la tienda por semana (M. E. Goldberg, G. J. Gorn, L. A. Peracchio y G. Bamossy, "Understanding Materialism Among Youth", *Journal of Consumer*

*Psychology*, 2003, 13(3):278-288). Suponga que quiere demostrar que los niños de su ciudad promedian más de cinco salidas a la tienda por semana. Sea  $\mu$  la media poblacional del número de veces que los niños de su ciudad salen a la tienda.

- Determine las hipótesis nula y alternativa.
- Explique el significado de los errores tipo I y tipo II en el contexto del escenario anterior.
- Suponga que realiza un estudio en la ciudad donde vive. Con base en estudios previos, usted supone que la desviación estándar del número de salidas a la tienda es de 1.6. Toma una muestra de 100 niños y descubre que el número medio de salidas a la tienda es de 5.47. Con un nivel de significancia de 0.01, ¿existen evidencias de que el número medio poblacional de salidas a la tienda es mayor que cinco por semana?
- Interprete el significado del valor-*p* en el inciso *c*).

**9.47** Las políticas de una sucursal bancaria específica establecen que sus cajeros automáticos deben contener efectivo suficiente para satisfacer a los clientes que hacen retiros durante todo el fin de semana. La aceptación del cliente depende de que tales servicios satisfagan sus necesidades. En esta sucursal, la cantidad media poblacional de dinero retirado del cajero automático por transacción durante el fin de semana es de 160 dólares, con una desviación estándar poblacional de 30 dólares. Suponga que en una muestra de 36 transacciones, se descubre que la cantidad media muestral de dinero retirado es de 172 dólares.

- Utilizando el método del valor crítico para probar la hipótesis, con un nivel de significancia de 0.05, ¿existen evidencias para creer que la cantidad media poblacional retirada es mayor que 160 dólares?
- Utilizando el método del valor-*p* para probar la hipótesis, con un nivel de significancia de 0.05, ¿existen evidencias para creer que la cantidad media poblacional retirada es mayor que 160 dólares?
- Interprete el significado del valor-*p* en este problema.
- Compare sus conclusiones para los incisos *a*) y *b*).

## 9.4 PRUEBA *t* DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ DESCONOCIDA)

En la mayoría de las situaciones de prueba de hipótesis que implican datos numéricos, no se conoce la desviación estándar poblacional  $\sigma$ . En su lugar, se utiliza la desviación estándar muestral  $S$ . Si se supone que la población tiene una distribución normal, la distribución muestral de la media seguirá una distribución *t* con  $n - 1$  grados de libertad. Si la población no tiene una distribución normal, todavía es válido utilizar la prueba *t* si el tamaño de la muestra es lo bastante grande como para que tenga efecto el teorema de límite central (vea la sección 7.2). La ecuación (9.2) define al estadístico prueba *t* para determinar la diferencia que existe entre la media muestral  $\bar{X}$  y la media poblacional  $\mu$  cuando se utiliza la desviación estándar muestral  $S$ .

### PRUEBA *t* DE HIPÓTESIS PARA LA MEDIA ( $\sigma$ DESCONOCIDA)

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \quad (9.2)$$

donde el estadístico de prueba *t* sigue una distribución *t* que tiene  $n - 1$  grados de libertad.

Para exemplificar el uso de esta prueba  $t$ , volvamos al escenario “Uso de la estadística” referente a la empresa de remodelaciones Saxon de la página 238. Durante los últimos cinco años, el monto medio por factura es de 120 dólares. Como contador de la empresa, si esta cantidad cambia, usted debe informar al departamento de finanzas. En otras palabras, se utiliza la prueba de hipótesis para tratar de demostrar que el monto medio por factura está aumentando o disminuyendo.

## Método del valor crítico

Para realizar esta prueba de hipótesis de dos colas, se utiliza el método de seis pasos que se encuentra en la exposición (9.1) de la página 280.

**Paso 1:**  $H_0: \mu = \$120$   
 $H_1: \mu \neq \$120$

La hipótesis alternativa contiene la afirmación que se trata de comprobar. Si se rechaza la hipótesis nula, se tendrán evidencias estadísticas de que la cantidad media por factura no supera 120 dólares. Si la conclusión estadística es “no rechazar  $H_0$ ”, entonces se concluirá que existen evidencias suficientes para demostrar que la cantidad media es distinta de la media a largo plazo de 120 dólares.

**Paso 2:** Usted seleccionó un tamaño de la muestra de  $n = 12$ . Decide utilizar  $\alpha = 0.05$ .

**Paso 3:** Puesto que no conoce  $\sigma$ , para este ejemplo utiliza la distribución  $t$  y el estadístico de prueba  $t$ . Debe suponerse que la población de facturas tiene una distribución normal. Esta suposición se analiza en la página 293.

**Paso 4:** Para una muestra de tamaño  $n$  dada, el estadístico de prueba  $t$  sigue una distribución  $t$  con  $n - 1$  grados de libertad. En la tabla E.3 se encuentran los valores críticos de la distribución  $t$  con  $12 - 1 = 11$  grados de libertad, como se ilustra en la figura 9.9 y en la tabla 9.3. Puesto que la hipótesis alternativa  $H_1$  de que  $\mu \neq 120$  dólares es *no direccional*, el área en la región de rechazo de la cola izquierda (inferior) de la distribución  $t$  es 0.025, y el área en la región de rechazo de la cola derecha (superior) de la distribución  $t$  es 0.025.

A partir de la tabla  $t$  como se presenta en la tabla E.3, que se muestra parcialmente en la tabla 9.3, los valores críticos son  $\pm 2.2010$ . La regla de decisión es:

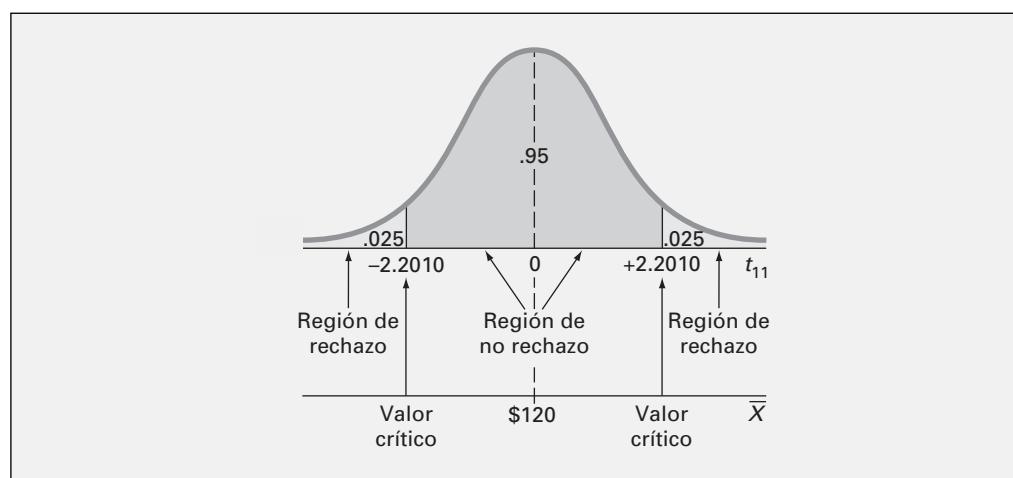
Rechazar  $H_0$  si  $t < -t_{11} = -2.2010$

o si  $t > t_{11} = +2.2010$ ,

de cualquier otra manera, no rechazar  $H_0$ .

**FIGURA 9.9**

Prueba de una hipótesis con respecto a la media ( $\sigma$  desconocida), con un nivel de significancia de 0.05 y con 11 grados de libertad.



**TABLA 9.3**

Determinación del valor crítico para un área de 0.025 en cada cola, con 11 grados de libertad, a partir de la tabla  $t$ .

Grados de libertad	Áreas de cola superior					
	.25	.10	.05	.025	.01	.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058

Fuente: Sintetizado de la tabla E.3.

**Paso 5:** Los siguientes datos, INVOICES, son las cantidades (en dólares) en una muestra de 12 facturas.

108.98 152.22 111.45 110.59 127.46 107.26  
93.32 91.97 111.56 75.71 128.58 135.11

Al utilizar las ecuaciones (3.1) y (3.10) de las páginas 73 y 82, o el resultado de Excel de la figura 9.10 o el de Minitab de la figura 9.11:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \$112.85 \quad \text{y} \quad S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \$20.80$$

A partir de la ecuación (9.2) de la página 290,

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{112.85 - 120}{\frac{20.80}{\sqrt{12}}} = -1.19$$

**FIGURA 9.10**

Hoja de Excel para la prueba  $t$  de una muestra de las facturas.

A	B
1 Mean Amount Per Invoice Hypothesis	
2	
3 Data	
4 Null Hypothesis $\mu =$	120
5 Level of Significance	0.05
6 Sample Size	12
7 Sample Mean	112.85
8 Sample Standard Deviation	20.8
9	
10 Intermediate Calculations	
11 Standard Error of the Mean	5.0044
12 Degrees of Freedom	11
13 $t$ Test Statistic	-1.1908
14	
15 Two-Tail Test	
16 Lower Critical Value	2.2010
17 Upper Critical Value	2.2010
18 $p$ -Value	0.2588
19 Do not reject the null hypothesis	

=B8/RAIZ(B6)  
=B6 - 1  
=(B7 - B4)/B11  
  
=-(DISTR.T.INV.(B5, B12))  
=DISTR.T.INV.(B5, B12)  
=DISTR.T(ABS(B13), B12, 2)

**FIGURA 9.11**

Resultado de Minitab para la prueba  $t$  de una muestra de las facturas.

Test of $\mu = 120$ vs not = 120							
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Amount	12	112.851	20.798	6.004	(99.636, 126.065)	-1.19	0.259

**Paso 6:** Puesto que  $-2.201 < t = -1.19 < 2.201$ , no se rechaza la hipótesis nula  $H_0$ . No existen evidencias suficientes para concluir que la cantidad media por factura no supera 120 dólares. Debe informarse al departamento de finanzas que la auditoría sugiere que el monto medio por factura no ha cambiado.

### Método del valor- $p$

**Paso 1-3:** Estos pasos se realizan exactamente igual que en el método del valor crítico.

**Paso 4:**  $t = -1.19$  (vea el paso 5 del método del valor crítico).

**Paso 5:** La hoja de Excel de la figura 9.10 y el resultado de Minitab de la figura 9.11 expresan que el valor- $p$  para esta prueba de dos colas es 0.259. Puesto que un valor- $p$  de 0.259 es mayor que  $\alpha = 0.05$ , no se rechaza  $H_0$ . Los datos no aportan evidencias suficientes para concluir que el monto medio por factura es distinto de 120 dólares. Debe informarse al departamento de finanzas que la auditoría sugiere que la cantidad media por factura no ha cambiado. El valor- $p$  indica que si la hipótesis nula fuera cierta, la probabilidad de que una muestra de 12 facturas pudiera tener una media mensual que variara por 7.15 dólares o más de los 120 dólares establecidos es de 0.259. En otras palabras, si el monto medio por factura es realmente de 120 dólares, entonces hay un 25.9% de posibilidades de observar una media muestral inferior a 112.85 dólares o superior a 127.15 dólares.

En el ejemplo anterior, resulta incorrecto decir que hay un 25.9% de posibilidades de que la hipótesis nula sea cierta. En ocasiones, quienes no están instruidos de manera adecuada en estadística utilizan esta mala interpretación del valor- $p$ . Recuerde que el valor- $p$  es una probabilidad condicional, que se calcula *suponiendo* que la hipótesis nula es cierta. En general, es apropiado asentar lo siguiente. Si la hipótesis nula es cierta, entonces hay una posibilidad de  $(\text{valor-}p)^*100\%$  de observar un resultado muestral por lo menos tan opuesto a la hipótesis nula como el resultado observado.

### Revisión de suposiciones

La prueba  $t$  de una muestra se utiliza cuando no se conoce la desviación estándar  $\sigma$ , y se estima empleando la desviación estándar muestral<sup>1</sup>  $S$ . La prueba  $t$  se considera un procedimiento *paramétrico clásico*, que hace varias suposiciones rigurosas que se deben sostener para garantizar que sus resultados sean válidos.

Para utilizar la prueba  $t$  de una muestra, se supone que los datos representan una muestra aleatoria procedente de una población con distribución normal. En la práctica, mientras el tamaño de la muestra no sea muy pequeño y la población no sea muy asimétrica, la distribución  $t$  brinda una buena aproximación de la distribución muestral de la media, cuando  $\sigma$  no se conoce.

Existen varias formas de evaluar la suposición de normalidad necesaria para utilizar la prueba  $t$ . Se puede observar qué tanto concuerdan los estadísticos muestrales con las propiedades teóricas de la distribución normal. También se podría usar un histograma, una gráfica de tallo y hoja, una gráfica de caja y bigote o una gráfica de probabilidad normal. Vea más detalles sobre la evaluación de la normalidad en la sección 6.3, página 194.

<sup>1</sup> Cuando se cuenta con un tamaño de muestra grande,  $S$  estima  $\sigma$  con precisión suficiente para que exista una poca diferencia entre las distribuciones  $t$  y  $Z$ . Por lo tanto, cuando el tamaño de la muestra es mayor que 120, se utiliza una prueba  $Z$  en lugar de una prueba  $t$ .

La figura 9.12 muestra el resultado de Excel para los datos de estadística descriptiva. La figura 9.13 muestra una gráfica de caja y bigote de Minitab. La figura 9.14 muestra una gráfica de probabilidad normal de Minitab.

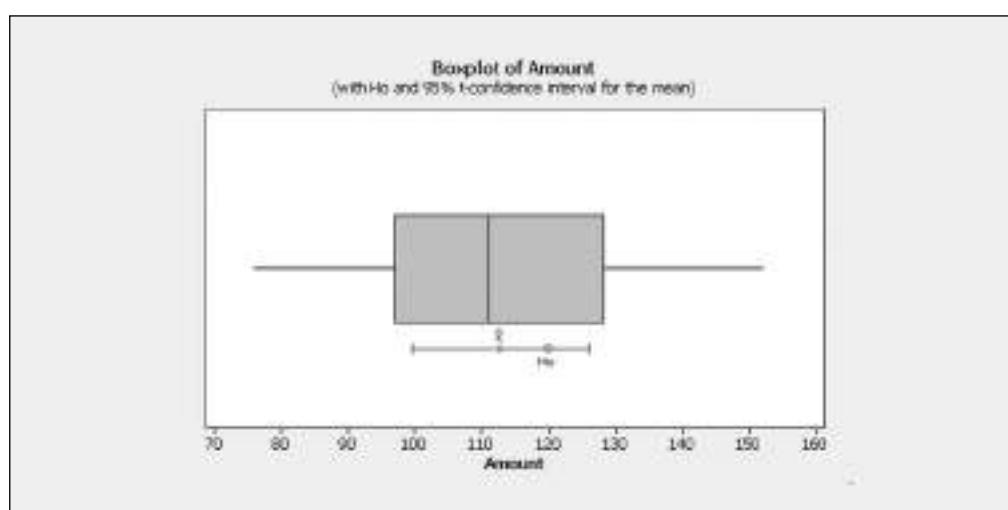
**FIGURA 9.12**

Estadística descriptiva para los datos de las facturas de venta.

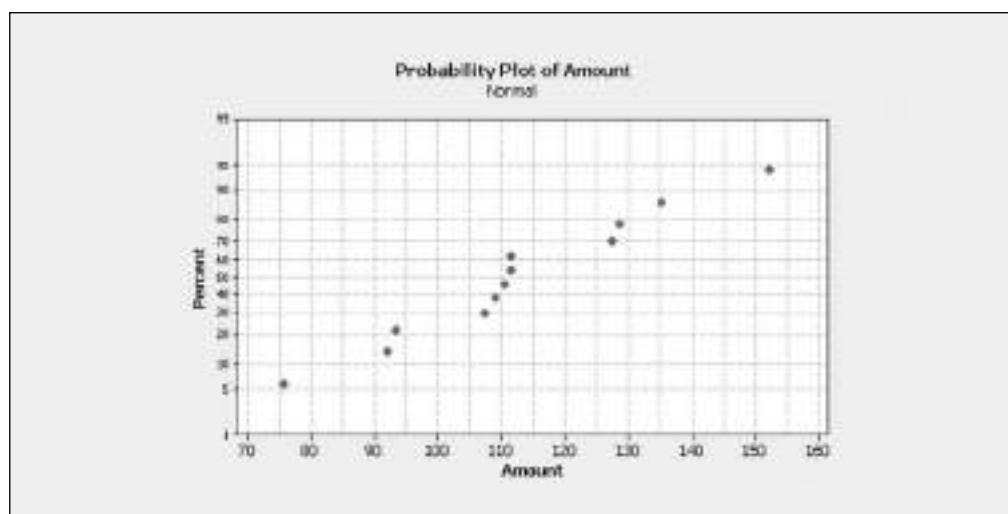
	A	B
1		<i>Invoice Amount</i>
2		
3	Mean	112.8508333
4	Standard Error	6.003863082
5	Median	111.82
6	Mode	#N/A
7	Standard Deviation	20.7979918
8	Sample Variance	432.5564629
9	Kurtosis	0.172707598
10	Skewness	0.133638902
11	Range	76.51
12	Minimum	75.71
13	Maximum	152.22
14	Sum	1354.21
15	Count	12
16	Largest(1)	152.22
17	Smallest(1)	75.71

**FIGURA 9.13**

Gráfica de caja y bigote de Minitab para los datos de las facturas de venta.

**FIGURA 9.14**

Gráfica de probabilidad normal de Minitab para los datos de las facturas de venta.



Puesto que la media es muy cercana a la mediana, los puntos sobre la gráfica de probabilidad normal parecen aumentar siguiendo una línea aproximadamente recta, y la gráfica de caja y bigote parece aproximadamente simétrica. Se puede suponer que la población de facturas tiene una distribución normal. Entonces, la suposición de normalidad es válida y también lo son los resultados del auditor.

La prueba  $t$  es una prueba **robusta**. No pierde fuerza si la forma de la población se desvía un poco de una distribución normal, en especial cuando el tamaño de la muestra es lo bastante grande como para permitir que el estadístico de prueba  $t$  reciba la influencia del teorema de límite central (vea la sección 7.2). Sin embargo, es probable que se obtengan conclusiones erróneas y que se pierda potencia estadística si la prueba  $t$  se utiliza de manera incorrecta. Si el tamaño de la muestra  $n$  es pequeño (es decir, menor que 30) y no se hace con facilidad la suposición de que la población subyacente tiene una distribución por lo menos aproximadamente normal, son más apropiados otros procedimientos de prueba *no paramétricos* (vea las referencias 1 y 2).

## PROBLEMAS PARA LA SECCIÓN 9.4

### Aprendizaje básico

ASISTENCIA  
de PH Grade

**9.48** Si en una muestra de  $n = 16$  seleccionada a partir de una población normal,  $\bar{X} = 56$  y  $S = 12$ , ¿cuál es el valor del estadístico de prueba  $t$  si usted está probando la hipótesis nula,  $H_0: \mu = 50$ ?

**9.49** En el problema 9.48, ¿cuántos grados de libertad hay en la prueba  $t$  de una muestra?

**9.50** En los problemas 9.48 y 9.49, ¿cuáles son los valores críticos de la tabla  $t$  si el nivel de significancia  $\alpha = 0.05$  y la hipótesis alternativa  $H_1$  es:

- a.  $\mu \neq 50$ ?
- b.  $\mu > 50$ ?

**9.51** En los problemas 9.48, 9.49 y 9.50, ¿cuál es su decisión estadística si la hipótesis alternativa  $H_1$  es:

- a.  $\mu \neq 50$ ?
- b.  $\mu > 50$ ?

**9.52** Si en una muestra de  $n = 16$  seleccionada a partir de una población sesgada a la izquierda,  $\bar{X} = 65$  y  $S = 21$ , ¿utilizaría la prueba  $t$  para probar la hipótesis nula,  $H_0: \mu = 60$ ? Discútalo.

**9.53** Si en una muestra de  $n = 160$  seleccionada a partir de una población sesgada a la izquierda,  $\bar{X} = 65$  y  $S = 21$ , ¿utilizaría la prueba  $t$  para probar la hipótesis nula,  $H_0: \mu = 60$ ? Discútalo.

### Aplicación de conceptos

Puede resolver los problemas 9.54 a 9.58 manualmente o con Excel, Minitab o SPSS. Le recomendamos resolver los problemas 9.59 a 9.65 con Excel, Minitab o SPSS.

AUTO  
Examen

**9.54** El director de admisión de una gran universidad advierte a los padres de los alumnos de primer ingreso sobre el costo de los libros durante un semestre típico. Seleccionó una muestra de 100 alumnos y registró sus gastos en libros durante el semestre. Luego calculó un costo medio muestral de 315.40 dólares y una desviación estándar de 43.2 dólares.

a. Utilizando un nivel de significancia de 0.10, ¿existe evidencia de que la media poblacional esté por arriba de 300 dólares?

b. ¿Cuál sería su respuesta al inciso a), si la desviación estándar fuera de 75 dólares y el nivel de significancia de 0.05?

c. ¿Cuál sería su respuesta al inciso a), si la media muestral fuera 305.11 dólares y la desviación estándar muestral de 43.20 dólares?

**9.55** En un artículo (Nanci Hellmich, “‘Supermarket Guru’ Has a Simple Mantra”, USA Today, 19 de junio, 2002, 70) se afirmó que la media de una visita típica al supermercado es de 22 minutos. Suponiendo que pretende probar dicha afirmación, usted selecciona una muestra de 50 compradores en el supermercado local. El tiempo de compras medio para la muestra de 50 compradores fue de 25.36 minutos, con una desviación estándar de 7.24 minutos. Utilizando un nivel de significancia de 0.10, ¿existen evidencias de que el tiempo de compras medio en el supermercado local es distinto al valor de 22 minutos que se afirma?

**9.56** Usted es gerente de un restaurante de comida rápida. Durante el mes pasado, el tiempo medio de espera en la ventanilla de servicio en el automóvil, medido a partir del momento en que el cliente realizó su pedido hasta que lo recibió, fue de 3.7 minutos. El dueño de la franquicia le ayuda a establecer un nuevo proceso que pretende reducir el tiempo de espera. Usted selecciona una muestra aleatoria de 64 pedidos. La media muestral del tiempo de espera es de 3.57 minutos, con una desviación estándar muestral de 0.8 minutos. Utilizando un nivel de significancia de 0.05, ¿existen evidencias de que la media poblacional del tiempo de espera es ahora menor que 3.7 minutos?

**9.57** Un fabricante de dulces de chocolate utiliza máquinas para empacar los dulces conforme pasan por una línea de llenado. A pesar de que los paquetes están etiquetados con un contenido de 8 onzas, la empresa quiere que tengan 8.17 onzas, de tal manera que virtualmente ninguno de los paquetes tenga menos de 8 onzas. De forma periódica se selecciona una muestra de 50 paquetes y, si se encuentran evidencias de que la cantidad media suministrada es distinta de 8.17 onzas, se detiene el proceso

de empacado. Suponga que la cantidad media abastecida en una muestra en particular de 50 paquetes es de 8.159 onzas, con una desviación estándar muestral de 0.051.

- a. ¿Existe evidencia de que la cantidad media poblacional es diferente de 8.17 onzas? (Utilice un nivel de significancia de 0.05.)

b. Calcule el valor-*p* e interprete su significado.

**9.58** Un fabricante de baterías para flash fotográfico tomó una muestra de 13 baterías, **BATTERIES**, de la producción diaria y las utilizó de manera continua hasta agotarlas. La vida en horas de las baterías hasta agotarse fue:

342	426	317	545	264	451	1,049
631	512	266	492	562	298	

- a. Con un nivel de significancia de 0.05, ¿existe evidencia de que la vida media de las baterías es mayor que 400 horas?
- b. Determine el valor-*p* en el inciso a) e interprete su significado.
- c. Utilizando la información anterior, ¿qué advertencia haría usted si el fabricante quisiera decir en sus anuncios que las baterías “durán más de 400 horas”?
- d. Suponga que el primer valor fuese 1,342 en lugar de 342, y repita los incisos a) a c), utilizando este valor. Comente sobre la diferencia en los resultados.

**9.59** En el estado de Nueva York, las cajas de ahorro tienen permitido vender cierta clase de seguro de vida, llamado Savings Bank Life Insurance (SBLI, Seguro de Vida de Caja de Ahorro). El proceso de aprobación consiste en una etapa de suscripción, la cual incluye una revisión de la solicitud, una visita a la oficina de información médica, posibles peticiones de información médica adicional y realización de exámenes médicos, así como la etapa de consolidación, durante la cual se generan las páginas de la póliza y se envían al banco para su entrega. La capacidad para entregar a los clientes oportunamente las pólizas aprobadas resulta vital para que este servicio sea rentable para el banco. En el transcurso de un mes, se seleccionó una muestra aleatoria de 27 pólizas aprobadas **INSURANCE** y se registró el siguiente tiempo de procesamiento total, en días:

73	19	16	64	28	28	31	90	60	56	31	56	22	18
45	48	17	17	17	91	92	63	50	51	69	16	17	

- a. En el pasado, el tiempo de procesamiento promedio fue de 45 días. Con un nivel de significancia de 0.05, ¿existe evidencia de que el tiempo de procesamiento medio ha variado de 45 días?
- b. ¿Qué suposición sobre la distribución de la población es necesario hacer en el inciso a)?
- c. ¿Cree que la suposición mencionada en el inciso b) se infringe seriamente? Explique por qué.

**9.60** Los siguientes datos representan la cantidad de bebida gaseosa envasada en una muestra de 50 botellas de 2 litros, de manera consecutiva. **DRINK** Los resultados se listan en forma horizontal en el orden de llenado:

2.109	2.086	2.066	2.075	2.065	2.057	2.052	2.044	2.036	2.038
2.031	2.029	2.025	2.029	2.023	2.020	2.015	2.014	2.013	2.014
2.012	2.012	2.012	2.010	2.005	2.003	1.999	1.996	1.997	1.992
1.994	1.986	1.984	1.981	1.973	1.975	1.971	1.969	1.966	1.967
1.963	1.957	1.951	1.951	1.947	1.941	1.941	1.938	1.908	1.894

- a. Con un nivel de significancia de 0.05, ¿existe evidencia de que la cantidad media de bebida gaseosa vertida en las botellas es distinta de 2.0 litros?
- b. Determine el valor-*p* en el inciso a) e interprete su significado.
- c. Evalúe de manera gráfica la suposición que hizo en el inciso a). ¿Son válidos los resultados de a)? ¿Por qué?
- d. Examine los valores de 50 botellas en su orden secuencial como se muestran en el planteamiento. ¿Existe algún patrón para los resultados? De ser así, ¿qué impacto podría tener ese patrón sobre la validez de los resultados del inciso a)?

**9.61** Una de las principales medidas de la calidad del servicio que brinda cualquier organización es la velocidad con la que responde a las quejas del cliente. Una gran tienda departamental propiedad de una familia, que vende muebles y pisos, incluyendo alfombras, emprendió una importante expansión durante los últimos años. En particular, el departamento de pisos se amplió de 2 equipos de instalación a un supervisor de instalación, un medidor y 15 equipos de instalación. El año pasado hubo 50 quejas relacionadas con la instalación de alfombras. Los siguientes datos representan el número de días transcurridos desde que se recibió la queja y su solución. **FURNITURE**

54	5	35	137	31	27	152	2	123	81	74	27
11	19	126	110	110	29	61	35	94	31	26	5
12	4	165	32	29	28	29	26	25	1	14	13
13	10	5	27	4	52	30	22	36	26	20	23
33	68										

- a. El supervisor de instalación asegura que el número medio de días transcurridos entre la recepción de la queja y su solución es de 20 días o menos. Con un nivel de significancia de 0.05, ¿existen evidencias de que tal afirmación no sea cierta (es decir, que el número medio de días sea mayor que 20)?
- b. ¿Qué suposición sobre la distribución de la población debe hacerse en el inciso a)?
- c. ¿Cree que la suposición mencionada en el inciso b) se infringe seriamente? Explique su respuesta.
- d. ¿Qué impacto podría tener su conclusión del inciso c) sobre la validez de los resultados del inciso a)?

**9.62** En un artículo publicado en *Quality Engineering*, se examina la viscosidad (resistencia al flujo) de un producto químico que se produce en lotes. Los datos correspondientes a 120 lotes se encuentran en el archivo **CHEMICAL**.

Fuente: Holmes y Mergen, “Parabolic Control Limits for the Exponentially Weighted Moving Average Control Charts”, *Quality Engineering*, 1992, 4(4): 487-495.

- a. En el pasado, la viscosidad media fue de 15.5. Con un nivel de significancia de 0.10, ¿existe evidencia de que la viscosidad media ha cambiado de 15.5?
- b. ¿Qué suposición sobre la distribución de la población debe hacerse en el inciso a)?
- c. ¿Cree que la suposición mencionada en el inciso b) se infringe seriamente? Explique su respuesta.

**9.63** Una función de un molino de acero consiste en cortar piezas de acero, formando piezas que se utilizan en el marco de los asientos delanteros de un automóvil. El acero se corta con una sierra de diamante y se requiere que las piezas resultantes tengan una longitud de  $\pm 0.005$  pulgadas con respecto a la especificada por la empresa automotriz. Los datos incluidos en el archivo **STEEL** corresponden a una muestra de 100 piezas de acero. La medición reportada es la diferencia, en pulgadas, que existe entre la longitud real de la pieza de acero, medida por medio de un dispositivo de rayo láser, y la longitud especificada para la pieza. Por ejemplo, un valor de  $-0.002$  corresponde a una pieza de acero que es 0.002 más corta que la longitud especificada.

- a. Con un nivel de significancia de 0.05, ¿existen evidencias de que la diferencia media no es igual a 0.0 pulgadas?
- b. Determine el valor- $p$  en el inciso a) e interprete su significado.
- c. ¿Qué suposición sobre las diferencias que existen entre la longitud real de la pieza de acero y la especificada debe hacerse en el inciso a)?
- d. Evalúe de manera gráfica la suposición mencionada en el inciso c). ¿Son válidos los resultados de a)? ¿Por qué?

**9.64** En el problema 3.55 de la página 112, se explica la operación de llenado de bolsas de té. Una importante característica de la calidad que resulta de interés en el proceso de llenado de bolsitas de té es el peso que contienen. Los datos del archivo **TEABAGS** son un arreglo ordenado del peso, en gramos, de una muestra compuesta por 50 bolsas de té producidas durante un turno de ocho horas.

- a. ¿Existe evidencia de que la cantidad media de té por bolsa es diferente de 5.5 gramos (utilice  $\alpha = 0.01$ )?
- b. Construya un intervalo de confianza estimado del 99% de la cantidad media poblacional de té por bolsa. Interprete este intervalo.
- c. Compare las conclusiones obtenidas en los incisos a) y b).

**9.65** La siguiente tabla contiene una muestra aleatoria de 30 fondos de inversión, tomada de los fondos de inversión men-

cionados en el *The Wall Street Journal* del 15 de junio de 2004. **CHANGE2004** La palabra “Cambio” representa el cambio (en dólares) en el valor del fondo al 14 de junio de 2004.

Fondos de inversión	Cambio
ABN AMRO Growth I	-0.22
Aim Funds HYld	0.00
Amer Advant Int Plan	-0.42
Artisan Funds SmCap	-0.24
Calif Trust S&P500	-0.23
Cohen and Steers Inst Rel	-0.55
Columbia Balance Z	-0.15
Delaware Large Cap Value A	-0.17
Dimension EmgMkt	-0.38
Dodge&Cox Stock	-1.50
Dreyfus It Inc	-0.09
Dreyfus OHMA	-0.02
Eaton Balanced	-0.06
Emerald Gr A	-0.18
Evergreen GLLeadA	-0.26
Federated Cap App	-0.20
Federated Intl Eq	-0.43
Fidelity Banking	-0.58
FirstAmerican LgCapValue	-0.17
Janus Balanced	-0.16
Kinetics Internet	-0.21
Merrill Lynch Equity Inc	-0.12
Nicholas Group Nich	-0.46
Northern Balanced A	-0.09
One Group DivMidCap	-0.24
Prudential Bear	0.03
Putnam Income	-0.02
South Trust Value	-0.14
Third Avenue Real Est	-0.26
Van Kampen Entc	-0.10

Fuente: *The Wall Street Journal*, 15 de junio, 2004.

- a. ¿Existe evidencia de que la media poblacional del valor del fondo cambia el 14 de junio de 2004? Utilice un nivel de significancia de 0.05.
- b. ¿Qué suposiciones hace para realizar la prueba en el inciso a)?
- c. Determine el valor- $p$  e interprete su significado.

## 9.5 PRUEBA Z DE HIPÓTESIS PARA LA PROPORCIÓN

En algunos casos, se requiere probar una hipótesis acerca de la proporción de una población  $\pi$  de valores ubicados dentro de una categoría específica, en vez de probar la media poblacional. Para empezar, se selecciona una muestra aleatoria y se calcula la **proporción de la muestra**,  $p = X/n$ . Luego se compara el valor de este estadístico con el valor del parámetro  $\pi$  establecido en la hipótesis, con el fin de decidir si se rechaza la hipótesis nula.

Si tanto el número de éxitos ( $X$ ) como el de fracasos ( $n - X$ ) son de por lo menos cinco cada uno, la distribución muestral de una proporción tiene una distribución muestral estandarizada aproximadamente normal. Para efectuar la prueba de hipótesis de la diferencia que existe entre la proporción muestral  $p$  y la proporción poblacional establecida en la hipótesis  $\pi$ , se utiliza la **prueba Z para la proporción** que muestra la ecuación (9.3).

## PRUEBA Z DE UNA MUESTRA, PARA LA PROPORCIÓN

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad (9.3)$$

donde

$$p = \frac{X}{n} = \frac{\text{número de éxitos en la muestra}}{\text{tamaño de la muestra}}$$

= proporción de éxitos en la muestra

$\pi$  = proporción hipotetizada de éxitos en la población

El estadístico de prueba Z mantiene una distribución normal estandarizada.

De manera alterna, al multiplicar numerador y denominador por  $n$ , se puede escribir el estadístico Z en términos del número de éxitos  $X$ , como muestra la ecuación (9.4).

## PRUEBA Z PARA LA PROPORCIÓN, EN TÉRMINOS DEL NÚMERO DE ÉXITOS

$$Z = \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \quad (9.4)$$

Utilicemos el siguiente estudio, publicado en *The Wall Street Journal*, para ilustrar la prueba Z de una muestra para una proporción. En ese estudio se planteó la pregunta: “¿Existe igual número de negocios caseros propiedad de hombres y mujeres?” El estudio de 899 negocios caseros reportó que 369 eran propiedad de mujeres (Eleena De Lisser y Dan Morse, “More Men Work at Home than Women, Study Shows”, *The Wall Street Journal*, 18 de mayo, 1999, B2).

Para este estudio, las hipótesis nula y alternativa se enunciaron de la siguiente manera:

$H_0: \pi = 0.50$  (es decir, la proporción de negocios caseros propiedad de mujeres es 0.50).

$H_1: \pi \neq 0.50$  (es decir, la proporción de negocios caseros propiedad de mujeres no es 0.50).

**Método del valor crítico**

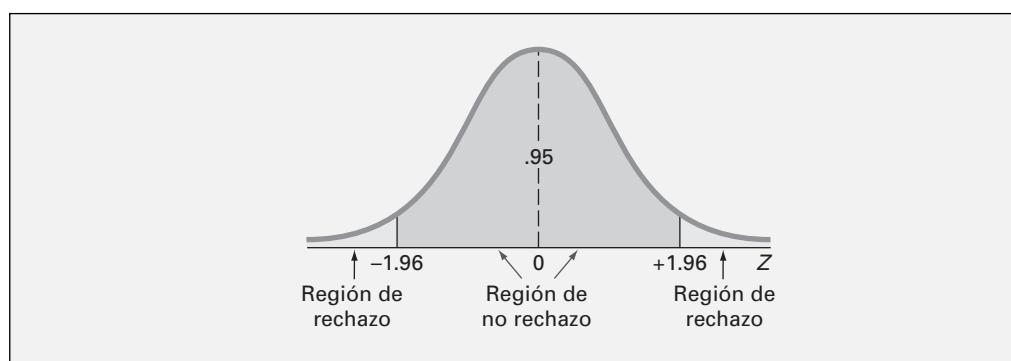
Puesto que se pretende conocer si la proporción de negocios caseros propiedad de mujeres es o no 0.50 (y la proporción de los que son propiedad de hombres es 0.50), utilice una prueba de dos colas. Si se selecciona un nivel de significancia de  $\alpha = 0.05$ , las regiones de rechazo y no rechazo quedan como muestra la figura 9.15, y la regla de decisión es:

Rechazar  $H_0$  si  $Z < -1.96$  o si  $Z > +1.96$ ;

de lo contrario, no rechazar  $H_0$ .

**FIGURA 9.15**

Prueba de hipótesis de dos colas para la proporción, con un nivel de significancia de 0.05.



Como 369 de los 899 negocios caseros son propiedad de mujeres,

$$p = \frac{369}{899} = 0.41046$$

Utilizando la ecuación (9.3),

$$Z \cong \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} = \frac{0.41046 - 0.50}{\sqrt{\frac{0.50(1 - 0.50)}{899}}} = \frac{-0.08954}{0.0167} = -5.37$$

o, utilizando la ecuación (9.4),

$$Z \cong \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} = \frac{369 - (899)(0.50)}{\sqrt{899(0.50)(0.50)}} = \frac{-80.5}{14.99} = -5.37$$

Como  $-5.37 < -1.96$ , se rechaza  $H_0$ . Se puede concluir que la proporción de negocios caseros propiedad de mujeres no es 0.50. La figura 9.16 muestra una hoja de trabajo de Excel para estos datos.

**FIGURA 9.16**

Hoja de trabajo de Excel para el estudio sobre propietarios de negocios caseros.

A	B
1	Ownership Proportion Hypothesis
2	
3	Data
4	Null Hypothesis $\pi = 0.5$
5	Level of Significance 0.05
6	Number of Successes 369
7	Sample Size 899
8	
9	Intermediate Calculations
10	Sample Proportion 0.4106
11	Standard Error 0.0167
12	Z Test Statistic -5.3697
13	
14	Two-Tail Test
15	Lower Critical Value -1.9600
16	Upper Critical Value 1.9600
17	$p$ Value 0.0000
18	Reject the null hypothesis

=B6/B7  
 =RAIZ(B4 \* (1 - B4)/B7)  
 =(B10 - B4)/B11  
  
 =DISTR. NORM. ESTAND. INV. (B5/2)  
 =DISTR. NORM. ESTAND. INV. (1-B5/2)  
 =2\*(1-DISTR. NORM. ESTAND.(ABS(B12)))

### Método del valor-*p*

Como método alternativo para tomar una decisión acerca de la prueba de hipótesis, se calcula el valor-*p*. Para esta prueba de dos colas, en la que la zona de rechazo se localiza en la cola inferior y la cola superior, es necesario encontrar el área por debajo de un valor *Z* de  $-5.37$  y por encima de un valor *Z* de  $+5.37$ . Las figuras 9.16 y 9.17 reportan un valor-*p* de 0.0000. Como este valor es menor que el nivel de significancia seleccionado ( $\alpha = 0.05$ ), se rechaza la hipótesis nula. Este valor-*p* tan pequeño indica que prácticamente no hay posibilidad de que la proporción muestral alcance un valor tan pequeño como 0.41046 si la proporción poblacional es 0.50.

**FIGURA 9.17**

Resultado de Minitab para el estudio sobre propietarios de negocios caseros.

Test of $p = 0.5$ vs $p \neq 0.5$						
Sample	X	N	Sample p	95% CI	Z-Value	P-Value
1	369	899	0.410456	(0.378300, 0.442612)	-5.37	0.000

**EJEMPLO 9.6****PRUEBA DE HIPÓTESIS PARA UNA PROPORCIÓN**

Una cadena de comida rápida acaba de desarrollar un novedoso proceso para asegurarse de que los pedidos de servicio en el automóvil se entreguen de manera correcta. Con el proceso anterior, los pedidos se entregaban correctamente el 88% de las veces. Se seleccionó una muestra de 100 pedidos despachados utilizando el nuevo proceso, de los cuales 92 se entregaron en forma correcta. Utilizando nivel de significancia de 0.01, ¿se puede concluir que el nuevo proceso ha aumentado la proporción de pedidos entregados correctamente?

**SOLUCIÓN** Las hipótesis nula y alternativa son:

$$H_0: \pi \leq 0.88 \text{ (es decir, la proporción de pedidos entregados correctamente es menor o igual que 0.88)}$$

$$H_1: \pi > 0.88 \text{ (es decir, la proporción de pedidos entregados correctamente es mayor que 0.88)}$$

Se utiliza la ecuación (9.3) de la página 298,

$$p = \frac{X}{n} = \frac{92}{100} = 0.92$$

$$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} = \frac{0.92 - 0.88}{\sqrt{\frac{0.88(1 - 0.88)}{100}}} = \frac{0.04}{0.0325} = 1.23$$

El valor- $p$  para  $Z > 1.23$  es 0.1093.

Utilizando el método del valor crítico,  $H_0$  se rechaza si  $Z > 2.33$ . Utilizando el método del valor- $p$ ,  $H_0$  se rechaza si el valor- $p < 0.01$ . Como  $Z = 1.23 < 2.33$  o el valor- $p = 0.1093 > 0.01$ , no se rechaza  $H_0$ . Se concluye que no hay evidencia suficiente de que el nuevo proceso haya aumentado la proporción de pedidos correctos por encima de 0.88.

**PROBLEMAS PARA LA SECCIÓN 9.5****Aprendizaje básico**

ASISTENCIA  
de PH Grade

**9.66** Si en una muestra aleatoria de 400 artículos, 88 están defectuosos, ¿cuál es la proporción muestral de artículos defectuosos?

ASISTENCIA  
de PH Grade

**9.67** En el problema 9.66, si la hipótesis nula establece que un 20% de los artículos están defectuosos, ¿cuál es el valor del estadístico de prueba  $Z$ ?

ASISTENCIA  
de PH Grade

**9.68** Suponga que en los problemas 9.66 y 9.67 usted está probando la hipótesis  $H_0: \pi = 0.20$  contra la hipótesis alternativa de dos colas  $H_1: \pi \neq 0.20$  y selecciona un nivel de significancia  $\alpha = 0.05$ . ¿Cuál es su decisión estadística?

**Aplicación de conceptos**

ASISTENCIA  
de PH Grade

AUTO  
Examen

**9.69** Un artículo publicado por *The Wall Street Journal* dice que más de la mitad de todos los estadounidenses preferirían recibir 100 dólares que un día libre en el trabajo. Esta afirmación se basa en una encuesta realizada por American Express Incentive Services, en la que 593 de 1,040 participantes dijeron

preferir los 100 dólares (Carlos Tejada, “Work Week”, *The Wall Street Journal*, 25 de julio, 2000, A1).

- Con un nivel de significancia de 0.05, ¿existe evidencia basada en los datos de la encuesta de que más de la mitad de todos los estadounidenses preferirían recibir 100 dólares que un día libre?
- Calcule el valor- $p$  e interprete su significado.

**9.70** Como resultado de la debilidad de la economía, se estima que durante 2003 sólo un 43% de los patrones estadounidenses contrató nuevos empleados. Pero hacia finales del año, la economía mostró signos de fortalecimiento. Según la encuesta realizada por la Society for Human Resource Management, 181 de 362 profesionales del manejo de recursos humanos planeaban contratar nuevos empleados en 2004 (Hane J. Kim, “Finally, 2004 May Be the Time to Seek a Raise”, *The Wall Street Journal*, 8 de enero, 2004, D4). Efectúe una prueba de hipótesis que trate de demostrar que la proporción de patrones que planearon contratar empleados nuevos en 2004 es mayor que la proporción de 0.43 correspondiente a 2003. Utilice el método de prueba de seis pasos y un nivel de significancia de 0.05.

**9.71** Un artículo publicado en el *Wall Street Journal* indica que los prejuicios con base en la edad se están convirtiendo en

un problema cada vez mayor en el mundo corporativo (Carol Hymowitz, “Top Executives Chase Youthful Appearance, But Miss Real Issue”, *The Wall Street Journal*, 17 de febrero, 2004, B1). En 2001, un estimado de 78% de los ejecutivos pensaban que los prejuicios con base en la edad eran un problema grave. En un estudio realizado en 2004 por ExecuNet, el 82% de los ejecutivos entrevistados consideraron los prejuicios con base en la edad como un problema grave. No se reveló el tamaño de la muestra utilizada en el estudio de 2004. Suponga que se entrevistó a 50 ejecutivos.

- Utilice el método de prueba de hipótesis de seis pasos, con un nivel de significancia de 0.05, para sondar y demostrar que la proporción de ejecutivos que creen que los prejuicios con base en la edad son un problema serio es mayor que el valor de 0.78 obtenido en 2001.
- Utilice los cinco pasos del método del valor-*p*. Interprete el significado del valor-*p*.
- Suponga ahora que el tamaño de la muestra fue de 1,000 y responda de nuevo a los incisos a) y b).
- Analice el efecto que tiene el tamaño de la muestra en el resultado de este análisis y, en general, la influencia del tamaño de la muestra en la prueba de hipótesis.

**9.72** En una encuesta realizada por el *Wall Street Journal* se preguntó a los participantes si creían en las calificaciones de eficiencia energética de automóviles y electrodomésticos; 552 respondieron que *sí*, y 531 respondieron que *no* (“What’s News Online”, *The Wall Street Journal*, 30 de marzo, 2004, D7).

- Utilice el método de prueba de hipótesis de seis pasos, con un nivel de significancia de 0.05, para explorar y demostrar que el porcentaje de personas que confían en las calificaciones de eficiencia energética es distinto de 50%.
- Utilice el método de cinco pasos del valor-*p*. Interprete el significado del valor-*p*.

**9.73** Uno de los principales desafíos que enfrentan los comercializadores por Internet es su capacidad para convertir a los visitantes en compradores (M. Totty, “Making the Sale”, *The Wall*

*Street Journal*, 24 de septiembre, 2001, R6). Esto se mide mediante la tasa de conversión, el porcentaje de visitantes que compran algo al visitar el sitio. Este artículo informó que la tasa de conversión de llbean.com fue del 10.1% y la de victoriasecret.com fue del 8.2%. Suponga que ambos sitios se rediseñaron con la finalidad de aumentar sus tasas de conversión. Se seleccionaron muestras de 200 visitantes en cada uno de los sitios rediseñados. Suponga que 24 visitantes de llbean.com y 25 de victoriasecret.com hicieron una compra.

- Con un nivel de significancia de 0.05, ¿existe evidencia de un aumento en la tasa de conversión de llbean.com?
- Con un nivel de significancia de 0.05, ¿existe evidencia de un aumento en la tasa de conversión de victoriasecret.com?

**9.74** Más que nunca antes, las mujeres profesionistas están renunciando a la maternidad, ante las limitantes de tiempo que implica su desarrollo profesional. Aún así, muchas mujeres todavía encuentran la manera de ascender en la escala corporativa y dedicar tiempo para tener hijos. Una encuesta realizada a 187 asistentes al encuentro *Las mujeres más poderosas en los negocios*, organizado por la revista *Fortune* en marzo de 2002, reveló que 133 tenían por lo menos un hijo (Carol Hymowitz, “Women Plotting Mix of Work and Family Won’t Find Perfect Plan”, *The Wall Street Journal*, 11 de junio, 2002, B1). Suponga que el grupo de 187 mujeres es una muestra de una población compuesta por todas las mujeres ejecutivas exitosas.

- ¿Cuál es la proporción muestral de mujeres ejecutivas exitosas que tienen hijos?
- Con un nivel de significancia de 0.05, ¿es posible afirmar que más de la mitad de todas las mujeres ejecutivas exitosas tienen hijos?
- Con un nivel de significancia de 0.05, ¿es posible afirmar que más de dos terceras partes de todas las mujeres ejecutivas exitosas tienen hijos?
- ¿Cree que la suposición de muestra aleatoria es válida? Explique su respuesta.



## 9.6 POSIBLES OBSTÁCULOS EN LAS PRUEBAS DE HIPÓTESIS Y CONSIDERACIONES ÉTICAS

Hasta aquí se han estudiado los conceptos fundamentales de las pruebas de hipótesis. Se utilizó la prueba de hipótesis para analizar las diferencias que existen entre estimaciones muestrales (es decir, estadísticos) y características poblacionales hipotéticas (es decir, parámetros), con el fin de tomar decisiones sobre las características subyacentes. El lector también aprendió a evaluar el riesgo implícito al tomar tales decisiones.

Mientras planea la realización de una prueba de hipótesis con base en una encuesta, investigación o experimento, debe plantear varias interrogantes que tienen por objeto garantizar el uso de la metodología apropiada. Durante la etapa de planeación, es necesario plantear y responder preguntas como las siguientes.

- ¿Cuál es el objetivo de la encuesta, estudio o experimento? ¿Cómo puede traducir ese objetivo en una hipótesis nula y una alternativa?
- ¿La prueba de hipótesis es una prueba de dos colas o una prueba de una cola?

3. ¿Puede seleccionar una muestra aleatoria a partir de toda la población de interés subyacente?
4. ¿Qué clases de datos recopilará de la muestra? ¿Las variables son numéricas o categóricas?
5. ¿Con qué nivel de significancia, o riesgo de cometer un error tipo I o tipo II, debe realizar la prueba de hipótesis?
6. ¿El tamaño de la muestra pretendido es lo suficientemente grande para alcanzar la potencia de la prueba correspondiente al nivel de significancia elegido?
7. ¿Cuál procedimiento de prueba estadística debe utilizar y por qué?
8. ¿Qué conclusiones e interpretaciones se obtendrán a partir de los resultados de la prueba de hipótesis?

Por lo anterior, es conveniente consultar a una persona con entrenamiento estadístico sustancial desde que inicia el proceso. Con mucha frecuencia, se consulta a esas personas de forma muy tardía en el proceso, después de recopilados los datos. Por lo general, todo lo que puede hacerse en una etapa tan avanzada es seleccionar el procedimiento de prueba estadístico que resulta mejor para los datos. El investigador se ve forzado a suponer que ciertos sesgos incorporados al estudio (causados por la mala planeación) son insignificantes. Pero ésta es una suposición enorme. La buena investigación implica una buena planeación. Para evitar los sesgos, desde el principio deben incluirse los controles adecuados.

Es necesario distinguir entre la metodología de investigación defectuosa y el comportamiento falso de ética. Las consideraciones éticas se presentan al manipular el proceso de prueba de hipótesis. Algunos de los problemas éticos que podrían surgir se relacionan con el método de recopilación de datos, el consentimiento informado de los sujetos sometidos a estudio, el tipo de prueba (una cola o dos colas), la elección del nivel de significancia  $\alpha$ , hurgar en los datos, la limpieza y eliminación de datos, y el reporte de los hallazgos.

**Método de recopilación de datos: aleatorización** Para eliminar la posibilidad de sesgos potenciales en los resultados, es necesario utilizar los métodos de recopilación de datos apropiados. Para obtener conclusiones significativas, los datos deben ser el resultado de una muestra aleatoria procedente de una población o un experimento en donde se utilizó el proceso de **aleatorización**. No debe permitirse que los participantes potenciales se autoselecciónen para un estudio, ni seleccionarlos a propósito. Además de los posibles problemas éticos que pudieran surgir, la falta de aleatorización provoca graves errores de cobertura o sesgos de selección que aniquilan la integridad del estudio.

**Consentimiento informado de los participantes humanos en “tratamiento”** Las consideraciones éticas exigen que todo individuo que en algún experimento va a ser sometido a algún “tratamiento” debe ser advertido del proyecto de investigación y de todos los efectos conductuales o físicos colaterales. El sujeto también debe proporcionar su consentimiento informado con respecto a su participación.

**Tipo de prueba: dos colas o una cola** Si se dispone de información previa que le conduzca a probar la hipótesis nula contra una alternativa específicamente dirigida, entonces una prueba de una cola es más potente que una de dos colas. Sin embargo, si sólo está interesado en las *diferencias* con respecto a la hipótesis nula, y no en la *dirección* de la diferencia, el procedimiento apropiado es la prueba de dos colas. Por ejemplo, si investigaciones y pruebas estadísticas realizadas previamente establecieron la diferencia en una dirección particular, o si una teoría científica establecida asienta que es posible que sólo se presenten resultados en una dirección, entonces es apropiada la prueba de una cola. Nunca es adecuado modificar la dirección de una prueba después de recopilados los datos.

**Selección del nivel de significancia  $\alpha$**  En un estudio bien diseñado, el nivel de significancia  $\alpha$  se selecciona antes de emprender la recopilación de los datos. No se debe alterar el nivel de signifi-

ficacia después del hecho para obtener un resultado específico. También es una buena costumbre reportar siempre el valor-*p*, y no sólo las conclusiones de la prueba de hipótesis.

**Hurgar en los datos** Nunca es permisible **hurgar en los datos**. No es ético realizar una prueba de hipótesis sobre un conjunto de datos, observar los resultados, y seleccionar después el nivel de significancia, o decidir entre una prueba de una o dos colas. Usted debe tomar esas decisiones antes de recabar los datos, con la finalidad de que sus conclusiones tengan sentido. En situaciones donde se consulta a un experto en estadística ya avanzado el proceso, con los datos ya en mano, resulta imperativo establecer las hipótesis nula y alternativa y seleccionar el nivel de significancia antes de efectuar la prueba de hipótesis. Además, no es válido modificar o eliminar de forma arbitraria los valores extremos o inusuales con el fin de alterar los resultados de las pruebas de hipótesis.

**Limpieza y eliminación de datos** La limpieza de datos no es igual que hurgar en ellos. En la etapa de preparación de datos, mientras los edita, codifica y transcribe, tendrá la oportunidad de revisarlos en busca de cualquier valor cuya medición parezca extrema o inusual. Tras revisar las observaciones anormales, debe construir una gráfica de tallo y hoja y/o una gráfica de caja y bigote, como preparación para una mayor presentación de datos y un análisis de confirmación. Esta etapa de análisis exploratorio de los datos brinda la oportunidad de limpiar el conjunto de éstos, mediante el marcado de los posibles atípicos, para verificarlos de nuevo contra los originales. Además, el análisis exploratorio de datos permite examinar los datos de manera gráfica con respecto a las suposiciones subyacentes al procedimiento de prueba de hipótesis específico.

El proceso de limpieza de datos genera un conflicto ético. En un estudio, ¿es correcto eliminar algún valor? La respuesta es un sí limitado. Si se puede determinar que una medición está incompleta o muy equivocada por un problema de equipo o por una situación conductual ajena al estudio, es válido eliminar ese valor. A veces no hay opción: tal vez un individuo abandone el estudio en el que participa, antes de que se efectúe la medición final. En un experimento o estudio bien diseñado, hay que decidir de antemano todas las normas relacionadas a la posible eliminación de datos.

**Reporte de los hallazgos** Al realizar investigación, habrá que documentar los resultados buenos y malos. Resulta inapropiado reportar los resultados de las pruebas de hipótesis que muestran significancia estadística, pero no los que no cuentan con evidencias suficientes dentro de los hallazgos. En caso de que no exista evidencia suficiente para rechazar  $H_0$ , debe aclararse que esto no demuestra que la hipótesis nula sea cierta. Lo que señala el resultado es que, con el tamaño de la muestra empleado, no existe información suficiente para *refutar* la hipótesis nula.

**Significancia estadística versus significancia práctica** Es necesario distinguir entre la existencia de un resultado estadísticamente significativo y su significancia práctica en el contexto de un campo de aplicación. A veces, gracias a un tamaño muy grande de la muestra, obtendrá un resultado estadísticamente significativo, pero con escasa significancia práctica. Por ejemplo, suponga que antes de una campaña nacional de comercialización basada en una serie de costosos anuncios en televisión, usted cree que la proporción de personas que reconocen su producto es 0.30. Al terminar la campaña, una encuesta aplicada a 20,000 personas indica que 6,168 reconocen su producto. Una prueba de una cola que pretende demostrar que la proporción es ahora mayor que 0.30 tiene como resultado un valor-*p* de 0.0047 y la conclusión estadística correcta es que aumentó la proporción de consumidores que ahora reconocen su producto. ¿Fue exitosa la campaña? El resultado de la prueba de hipótesis señala un aumento significativo estadísticamente en cuanto al reconocimiento de marca, pero ¿este aumento es importante en la práctica? Usted estima que ahora la proporción poblacional es de  $6,168/20,000 = 0.3084$  o 30.84%. Este aumento es menos de 1% mayor que el valor del 30% establecido en la hipótesis. ¿Los enormes gastos asociados con la campaña de comercialización generaron un resultado con un aumento significativo del reconocimiento de marca? Ante el escaso impacto que en el mundo real tiene un aumento inferior al 1% sobre la estrategia general de comercialización, y ante los enormes gastos relacionados con la campaña, debe concluirse que ésta no fue exitosa. Por otra parte, si la campaña aumentara el reconocimiento de marca en un 20%, usted podría concluir que fue exitosa.

En resumen, al analizar los aspectos éticos relacionados con las pruebas de hipótesis, la clave radica en *la intención*. Es indispensable distinguir entre un mal análisis de los datos y una práctica falta de ética. Esta última se presenta cuando, *intencionalmente*, los investigadores crean un sesgo de selección al recopilar los datos, manipulan el tratamiento de sujetos humanos sin su consentimiento informado, indagan en los datos para seleccionar el tipo de prueba (dos colas o una cola) y/o el nivel de significancia, ocultan los hechos eliminando los valores que no respaldan una hipótesis establecida, o no reportan hallazgos pertinentes.

## RESUMEN

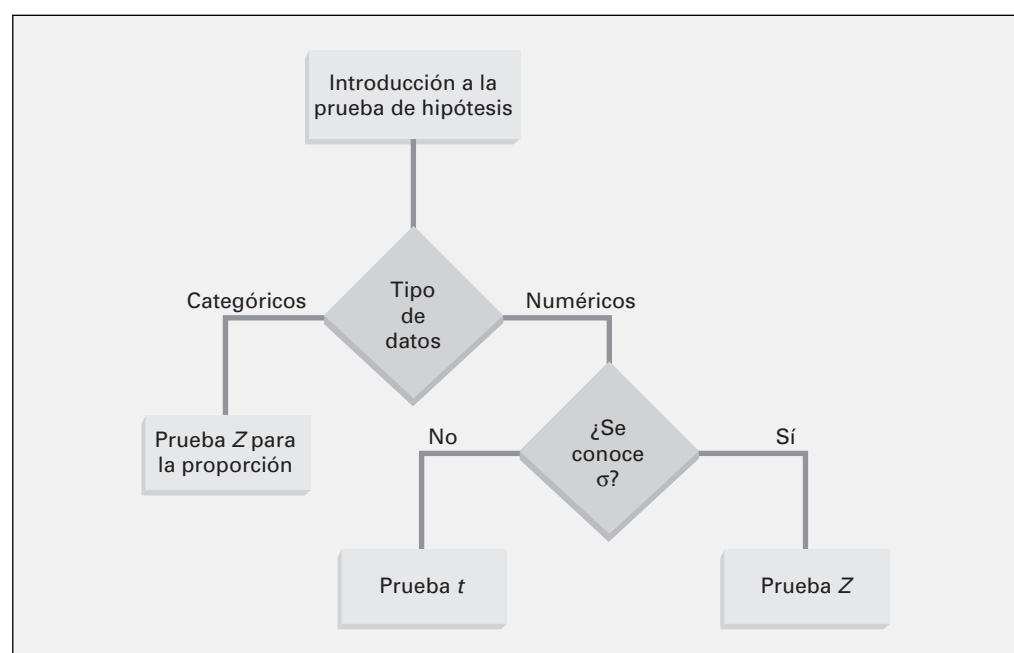
En este capítulo se presentaron los fundamentos de la prueba de hipótesis. El lector aprendió a realizar las pruebas  $Z$  y  $t$  en la media poblacional, una prueba  $Z$  para la proporción poblacional, y cómo utilizar las pruebas de hipótesis para vigilar y mejorar un proceso de llenado de cereal. Los dos capítulos si-

guientes se basan en los fundamentos de las pruebas de hipótesis aquí analizados.

En la figura 9.18 se presenta un mapa para seleccionar la correcta prueba de hipótesis de una muestra.

**FIGURA 9.18**

Mapa para seleccionar la prueba de una muestra.



Al decidir qué tipo de prueba se va a utilizar, hay que plantearse las siguientes preguntas:

- ¿La prueba implica a una variable numérica o categórica? Si la respuesta es una variable categórica, utilice la prueba  $Z$  para la proporción.
- Si la prueba implica a una variable numérica, ¿se conoce la desviación estándar poblacional? Si conoce la desviación

estándar poblacional, utilice la prueba  $Z$  para la media. Si no conoce la desviación estándar poblacional, utilice la prueba  $t$  para la media.

En la tabla 9.4 se encuentra una lista con los temas estudiados en este capítulo.

**TABLA 9.4**

Resumen de los temas incluidos en el capítulo 9.

Tipo de datos		
Tipo de análisis	Numéricos	Categóricos
Prueba de hipótesis relativa a un solo parámetro	Prueba $Z$ de hipótesis para la media ( <b>sección 9.2</b> ) Prueba $t$ de hipótesis para la media ( <b>sección 9.4</b> )	Prueba $Z$ de hipótesis para la proporción ( <b>sección 9.5</b> )

# FÓRMULAS IMPORTANTES

Prueba Z de hipótesis para la media ( $\alpha$  conocida)

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (9.1)$$

Prueba t de hipótesis para la media ( $\sigma$  desconocida)

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \quad (9.2)$$

Prueba Z de una muestra, para la proporción

$$Z \cong \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad (9.3)$$

Prueba Z para la proporción, en términos del número de éxitos

$$Z \cong \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \quad (9.4)$$

# CONCEPTOS CLAVE

aleatorización 302  
 coeficiente de confianza  $(1 - \alpha)$  275  
 error tipo I 275  
 error tipo II 275  
 estadístico de prueba 274  
 estadístico de prueba Z 278  
 hipótesis alternativa  $H_1$  273  
 hipótesis nula ( $H_0$ ) 272  
 hurgar en los datos 303

nivel de confianza 275  
 nivel de significancia ( $\alpha$ ) 275  
 potencia de una prueba estadística 276  
 proporción de la muestra 297  
 prueba de dos colas 278  
 prueba de hipótesis 272  
 prueba de una cola 286  
 prueba direccional 286  
 prueba t de hipótesis para la media 290

prueba Z de hipótesis para la media 278  
 prueba Z para la proporción 297  
 región de no rechazo 274  
 región de rechazo 274  
 riesgo  $\alpha$  276  
 valor crítico 275  
 valor- $p$  281  
 robusta 295  
 $\alpha$  (nivel de significancia) 275

# PROBLEMAS DE REPASO

## Revisión de su aprendizaje

- 9.75** ¿Cuál es la diferencia entre una hipótesis nula  $H_0$  y una hipótesis alternativa  $H_1$ ?
- 9.76** ¿Cuál es la diferencia entre el error tipo I y el error tipo II?
- 9.77** ¿Qué se quiere decir con potencia de una prueba?
- 9.78** ¿Cuál es la diferencia entre una prueba de una cola y una prueba de dos colas?
- 9.79** ¿Qué quiere decir por valor- $p$ ?
- 9.80** ¿Cómo un intervalo de confianza permite estimar la media poblacional, brindando conclusiones para la prueba de hipótesis correspondiente a la media poblacional?
- 9.81** ¿Cuál es el método del valor crítico de seis pasos para la prueba de hipótesis?
- 9.82** ¿Cuáles son algunos de los aspectos éticos relacionados con la realización de una prueba de hipótesis?

## Aplicación de conceptos

- 9.83** En un artículo publicado en *Marketing News* (Thomas T. Semon, “Consider a Statistical Insignificance Test”, *Marketing News*, 1 de febrero, 1999) se argumenta que con frecuencia el nivel de significancia utilizado al comparar dos productos es muy bajo; es decir, a veces se debe utilizar un valor  $\alpha$  mayor que 0.05. De manera específica, el artículo recalcó una prueba de proporción de clientes potenciales con preferencia por el producto 1 sobre el producto 2. La hipótesis nula establece que la proporción de la población de clientes potenciales que prefiere el producto 1 es 0.50, y la hipótesis alternativa establece que no es igual a 0.50. El valor- $p$  para la prueba es 0.22. El artículo sugiere que en algunos casos, esto debe ser evidencia suficiente para rechazar la hipótesis nula.

- Determine las hipótesis nula y alternativa en términos estadísticos.
- Explique los riesgos relacionados con los errores tipo I y tipo II.
- ¿Cuáles son las consecuencias de rechazar la hipótesis nula para un valor- $p$  de 0.22?

- d. ¿Por qué cree que el artículo sugiere aumentar el valor de  $\alpha$ ?
- e. ¿Usted qué haría en esta situación?
- f. ¿Cuál sería su respuesta al inciso e) si el valor- $p$  fuera igual a 0.12? ¿Y si fuera igual a 0.06?

**9.84** La Quinta Motor Inns desarrolló un modelo computarizado para auxiliar en el pronóstico de rentabilidad de los lugares que se están considerando como ubicación de hoteles nuevos. Si el modelo computarizado pronostica grandes ganancias, La Quinta adquiere el predio y construye un nuevo hotel. Si el modelo pronostica ganancias pequeñas o moderadas, La Quinta prefiere no operar en ese lugar (Sheryl E. Kimes y James A. Fitzsimmons, "Selecting Profitable Hotel Sites at La Quinta Motor Inns", *Interfaces*, vol. 20, marzo-abril de 1990, 12-20). Este procedimiento de toma de decisiones se puede expresar en el marco de la prueba de hipótesis. La hipótesis nula establece que el sitio no es una ubicación rentable. La hipótesis alternativa establece que el sitio es una ubicación rentable.

- a. Explique los riesgos relacionados con cometer un error tipo I.
- b. Explique los riesgos relacionados con cometer un error tipo II.
- c. ¿Qué tipo de error cree que están tratando de evitar a toda costa los ejecutivos de La Quinta Motor Inns? Explique su respuesta.
- d. ¿Cómo influyen los cambios en el criterio de rechazo sobre las probabilidades de cometer errores tipo I y II?

**9.85** Un estudio realizado en 1999 por la General Accounting Office (GAO) descubrió que alrededor de la tercera parte de los 23.4 millones de jubilados estadounidenses con 65 años de edad o más complementaban sus servicios médicos proporcionados por la seguridad social con algún tipo de cobertura por parte del patrón (Carlos Tejada, "Work Week", *The Wall Street Journal*, 26 de junio, 2002, B5). El artículo sugiere que esta proporción va en aumento. Suponga que en un estudio actual, una muestra compuesta por 500 jubilados de 65 años o más señala que 185 de ellos complementan los servicios médicos de seguridad social con alguna forma de cobertura por parte del patrón.

- a. Con un nivel de significancia de 0.01, ¿existen evidencias de que la proporción de jubilados con 65 años o más que complementan los servicios de salud con alguna forma de cobertura patronal sea ahora superior a la tercera parte?
- b. Calcule el valor- $p$  e interprete su significado.

**9.86** El propietario de una gasolinera desea estudiar los hábitos de consumo de gasolina de los conductores en su estación. Una semana cualquiera, se selecciona una muestra aleatoria de 60 conductores, con los siguientes resultados

- Cantidad adquirida:  $\bar{X} = 11.3$  galones,  $S = 31$  galones.
- 11 conductores compraron gasolina Premium.
- a. Con un nivel de significancia de 0.05, ¿existe evidencia de que la compra media es distinta de 10 galones?
- b. Encuentre el valor- $p$  en el inciso a).
- c. Con un nivel de significancia de 0.05, ¿existe evidencia de que menos del 20% de todos los conductores compran gasolina Premium en esa gasolinera?
- d. ¿Cuál sería su respuesta al inciso a) si la media muestral fuera de 10.3 galones?
- e. ¿Cuál sería su respuesta al inciso c) si fueran siete los conductores que compraron gasolina Premium?

**9.87** El auditor de una oficina gubernamental recibe la encuesta de evaluar los reembolsos por visitas de médicos a la oficina pagadas por la seguridad social. La auditoría se realiza sobre una muestra de 75 de los reembolsos, con los siguientes resultados:

- En 12 de las visitas a la oficina, se reembolsó una cantidad incorrecta.
- El monto de los reembolsos fue:  $\bar{X} = \$93.70$ ,  $S = \$34.55$ .
- a. Utilizando un nivel de significancia de 0.05, ¿existe evidencia de que el reembolso medio es menor que 100 dólares?
- b. Utilizando un nivel de significancia de 0.05, ¿existe evidencia de que la proporción de reembolsos incorrectos en la población es mayor que 0.10?
- c. Analice las suposiciones subyacentes de la prueba utilizada en el inciso a).
- d. ¿Cuál sería su respuesta al inciso a) si la media muestral fuera de 90 dólares?
- e. ¿Cuál sería su respuesta al inciso b) si fueran 15 las visitas a la oficina con reembolso incorrecto?

**9.88** Una sucursal bancaria ubicada en una zona comercial de la ciudad, desarrolló un proceso mejorado para atender a sus clientes durante la hora del almuerzo, del mediodía a la 1:00 PM. Durante una semana, se registra el tiempo de espera en minutos (definido como el tiempo transcurrido desde que el cliente se forma en la fila hasta que llega a la ventanilla del cajero) de todos los clientes durante ese horario. Se selecciona una muestra aleatoria de 15 clientes, y los resultados son los siguientes: **BANK1**

4.21 5.55 3.02 5.13 4.77 2.34 3.54 3.20  
4.50 6.10 0.38 5.12 6.46 6.19 3.79

- a. Utilizando un nivel de significancia de 0.05, ¿existen evidencias de que el tiempo medio de espera es menor que 5 minutos?
- b. ¿Qué suposición debe sostener con la finalidad de realizar la prueba del inciso a)?
- c. Evalúe esta suposición por medio de un método gráfico. Analícelo.
- d. Durante la hora del almuerzo, un cliente llega al banco y pregunta al gerente cuánto deberá esperar. El gerente de la sucursal responde: "Casi con toda seguridad, no más de 5 minutos". Con base en los resultados del inciso a), evalúe esta afirmación.

**9.89** Una empresa manufacturera fabrica aislantes eléctricos. Si los aislantes se rompen al estar en uso, es probable que ocurra un corto circuito. Para poner a prueba la fuerza de los aislantes, se efectúa una prueba de destrucción para determinar cuánta fuerza se necesita para romperlos. La fuerza se mide observando cuántas libras se aplican al aislante antes de que se rompa. A continuación se muestran los datos de 30 observaciones procedentes de este experimento: **FORCE**

*Fuerza (en número de libras necesarias para romper el aislante)*

1,870 1,728 1,656 1,610 1,634 1,784 1,522 1,696 1,592 1,662  
1,866 1,764 1,734 1,662 1,734 1,774 1,550 1,756 1,762 1,866  
1,820 1,744 1,788 1,688 1,810 1,752 1,680 1,810 1,652 1,736

- a. Utilizando un nivel de significancia de 0.05, ¿existen evidencias de que la fuerza media es mayor que 1,500 libras?
- b. ¿Qué suposición debe hacerse para realizar la prueba del inciso a)?
- c. Evalúe esta suposición por medio de un método gráfico. Analícela.
- d. Con base en su respuesta al inciso a), ¿qué se concluye sobre la resistencia de los aislantes?

**9.90** Una importante característica de calidad utilizada por el fabricante de las tejas de asfalto “Boston” y “Vermont” es la cantidad de humedad que contienen las tejas al momento de empacarlas. Los clientes tal vez crean que compraron un producto de mala calidad al encontrar humedad y tejas mojadas dentro del empaque. En algunos casos, la humedad excesiva provoca que se caigan los gránulos pegados a la teja para darle textura y color, lo que da como resultado problemas de apariencia. Para vigilar la cantidad de humedad presente, la empresa realiza pruebas. Se pesa una teja y luego se seca. Se pesa de nuevo la teja y, con base en la cantidad de humedad extraída del producto, se calculan las libras de humedad por cada 100 pies cuadrados. La empresa quisiera demostrar que el contenido medio de humedad es menor que 0.35 libras por 100 pies cuadrados. El archivo MOISTURE incluye 36 mediciones (en libras por cada 100 pies cuadrados) de las tejas “Boston” y 31 de las tejas “Vermont”.

- a. Para las tejas “Boston”, ¿existen evidencias, con un nivel de significancia de 0.05, de que el contenido medio de humedad es menor que 0.35 libras por cada 100 pies cuadrados?
- b. Interprete el significado del valor-*p* en el inciso a).
- c. Para las tejas “Vermont”, ¿existe evidencia con un nivel de significancia de 0.05, de que el contenido medio de humedad es menor que 0.35 libras por cada 100 pies cuadrados?
- d. Interprete el significado del valor-*p* en el inciso c).
- e. ¿Qué suposición debe hacerse para realizar las pruebas de los incisos a) y c)?
- f. Evalúe la suposición correspondiente a cada tipo de teja utilizando un método gráfico. Discútala.

**9.91** Los estudios realizados por el fabricante de las tejas de asfalto “Boston” y “Vermont” mostraron que el peso del producto es un factor determinante en la percepción de calidad del cliente. Además, el peso representa la cantidad de materias primas utilizadas y, por lo tanto, es muy importante para la empresa desde el punto de vista de los costos. En la última etapa del ensamblaje, las tejas se empacan antes de colocar los paquetes

sobre plataformas de madera. Una vez que se llena la plataforma (para casi todas las tejas, una plataforma sostiene 16 cuadros de tejas), se pesa y se registra esta medición. El archivo Pallet contiene el peso (en libras) de una muestra compuesta por 368 plataformas de tejas Boston y 330 plataformas de tejas Vermont.

- a. Para las tejas Boston, ¿existe evidencia de que el peso medio es distinto a 3,150 libras?
- b. Interprete el significado del valor-*p* en el inciso a).
- c. Para las tejas Vermont, ¿existe evidencia de que el peso medio es distinto a 3,700 libras?
- d. Interprete el significado del valor-*p* en el inciso c).
- e. ¿La suposición necesaria para los incisos a) y c) se infringe seriamente?

**9.92** El fabricante de las tejas de asfalto “Boston” y “Vermont” ofrece a sus clientes una garantía de 20 años en la mayoría de sus productos. Para determinar si una teja dura tanto como el periodo de garantía, en la planta se realiza una prueba de vida acelerada. En la prueba de vida acelerada, realizada en un laboratorio, la teja se expone a las tensiones que recibiría en toda su vida útil de uso normal, mediante un experimento que lleva tan sólo unos minutos. En esta prueba se cepilla repetidamente una teja durante un breve lapso, y se pesa la cantidad de gránulos (en gramos) desprendidos por el cepillado. Se espera que las tejas con menor desprendimiento duren más en uso normal que las que experimentan gran cantidad de desprendimiento. El archivo GRANULE contiene los datos de una muestra compuesta por 170 medidas realizadas en las tejas Boston de la empresa y 140 medidas realizadas en las tejas Vermont.

- a. Para las tejas Boston, ¿existe evidencia de que la pérdida media de gránulos es distinta de 0.05 gramos?
- b. Interprete el significado del valor-*p* en el inciso a).
- c. Para las tejas Vermont, ¿existe evidencia de que la pérdida media de gránulos es distinta de 0.05 gramos?
- d. Interprete el significado del valor-*p* en el inciso c).
- e. ¿Se infringe seriamente la suposición necesaria para responder los incisos a) y c)?

## Ejercicios de reporte escrito

**9.93** Consulte los resultados de los problemas 9.90 a 9.92, relacionados con las tejas “Boston” y “Vermont”, para escribir un reporte en el que se evalúe el nivel de humedad, peso y pérdida de gránulos para ambos tipos de teja.

## CASO ACTUAL

### ADMINISTRACIÓN DEL SPRINGVILLE HERALD

Continuando con la supervisión de la tonalidad oscura en la impresión del periódico, ya descrita en el caso *Administración del Springville Herald* del capítulo 6, el departamento de producción quiere asegurarse de que la tonalidad oscura media de la impresión de todos los ejemplares sea por lo menos de 0.97 sobre una escala estándar en la que el valor que se pretende es igual a 1.0. Se seleccionó una muestra aleatoria de 50 ejemplares y se midió la tonalidad oscura en cada uno de ellos. Calcule los estadísticos muestrales y determine si existen evidencias de que la tonalidad oscura media es menor que 0.97. Redacte un memorando para la gerencia en el que sintetice sus conclusiones.

Tonalidad oscura de 50 periódicos										
0.854	1.023	1.005	1.030	1.219	0.977	1.044	0.778	1.122	1.114	
1.091	1.086	1.141	0.931	0.723	0.934	1.060	1.047	0.800	0.889	
1.012	0.695	0.869	0.734	1.131	0.993	0.762	0.814	1.108	0.805	
1.223	1.024	0.884	0.799	0.870	0.898	0.621	0.818	1.113	1.286	
1.052	0.678	1.162	0.808	1.012	0.859	0.951	1.112	1.003	0.972	

SH9

## CASO WEB

*Aplique sus conocimientos sobre las pruebas de hipótesis a este Caso Web, que es continuación de la controversia sobre el llenado de paquetes de cereal del Caso Web del capítulo 7.*

Hace poco, Oxford Cereals realizó un experimento público en el que asegura haber desacreditado con éxito las afirmaciones de grupos como el TriCities Consumers Concerned About Cereal Companies That Cheat (TCCACCTC), que aseguran que Oxford Cereals engaña a los consumidores al empacar cereales con pesos inferiores a los especificados en la etiqueta. Analice el boletín de prensa de Oxford Cereals y los documentos de respaldo que describen el experimento, que se encuentran en el sitio Web de la empresa: [www.prenhall.com/Springville/OC\\_WinTrust.htm](http://www.prenhall.com/Springville/OC_WinTrust.htm) y luego responda lo siguiente:

1. ¿Son válidos los resultados de la prueba independiente? ¿Por qué? Si usted estuviera realizando el experimento, ¿hay algo que cambiaría?
2. ¿Los resultados respaldan la afirmación de que Oxford Cereals no está engañando a sus clientes?
3. ¿Resulta sorpresiva la afirmación del director general de Oxford acerca de que muchas cajas de cereal contienen más de 368 gramos? ¿Es cierto?
4. ¿Podría existir alguna circunstancia en la que ambos resultados, los del experimento público y los de la TCCACCTC, sean correctos? Explique su respuesta.

## REFERENCIAS

1. Bradley, J. V., *Distribution-Free Statistical Tests* (Englewood Cliffs, NJ: Prentice Hall, 1968).
2. Daniel, W., *Applied Nonparametric Statistics*, 2a. ed. (Boston, MA: Houghton Mifflin, 1990).
3. *Microsoft Excel 2003* (Redmond, WA: Microsoft Corporation, 2003).
4. *Minitab for Windows Version 14* (State College PA: Minitab, Inc., 2004).
5. *SPSS Base 12.0 Brief Guide* (Upper Saddle River, NJ: Prentice Hall, 2003).

# Apéndice 9 Uso del software para las pruebas de hipótesis con una muestra

## A9.1 EXCEL

### Para la prueba Z para la media, conociendo sigma

Abra el archivo **Z Mean.xls**, que se muestra en la figura 9.5 de la página 282. Esta hoja de trabajo ya contiene las entradas del ejemplo 9.2 relativas a la Oxford Cereal Company. Esta hoja utiliza las funciones NORMSINV y NORMSDIST (vea la sección G.16 para mayor información). Para adaptar esta hoja a otros problemas, modifique los valores de la hipótesis nula, el nivel de significancia, la desviación estándar de la población, el tamaño de la muestra y la media muestral que se encuentran en las celdas sombreadas de las filas 4 a 8.

*O* Vea la sección G.16 (**Prueba Z para la media, conociendo sigma**) si desea que PHStat2 genere una hoja de trabajo para usted.

### Para la prueba t para la media, conociendo sigma

Abra el archivo **T Mean.xls**, que se muestra en la figura 9.10 de la página 292. Esta página ya contiene las entradas del ejemplo sobre facturas de la sección 9.4. En esta hoja se utilizan las funciones TINV y TDIST. (Para mayor información, véase la sección G.17.) Para adaptar esta hoja de trabajo a otros problemas, modifique los valores de la hipótesis nula, el nivel de significancia, el tamaño de la muestra, la media muestral y la desviación estándar de la muestra que se encuentran en las celdas sombreadas de las filas 4 a 8.

*O* Vea la sección G.17 (**Prueba t para la media, conociendo sigma**) si desea que PHStat2 genere una hoja de trabajo para usted.

### Para la prueba Z para la proporción

Abra el archivo **Z Proportion.xls**, que se muestra en la figura 9.16 de la página 299. Esta hoja ya contiene las entradas del ejemplo sobre propietarios de negocios de la sección 9.5. Esta hoja de trabajo utiliza las funciones NORMSINV y NORMSDIST (para mayor información, véase la sección G.18). Para adaptar esta hoja de trabajo a otros problemas, modifique los valores de la hipótesis nula, el nivel de significancia, el número de éxitos y el tamaño de la muestra que se encuentran en las celdas sombreadas de las filas 4 a 7.

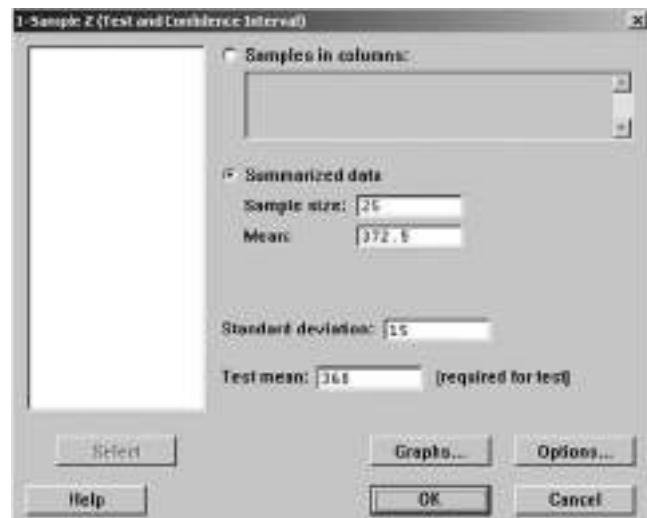
*O* Vea la sección G.18 (**Prueba Z para la proporción**) si desea que PHStat2 genere una hoja de trabajo para usted.

## A9.2 MINITAB

### Uso de Minitab para la prueba Z de hipótesis para la media ( $\sigma$ conocida)

Conociendo  $\sigma$ , puede utilizar Minitab para una prueba de la media. Para ilustrar la prueba de hipótesis para la media cuando se conoce  $\sigma$ , regrese al ejemplo de la Oxford Cereal Company analizado en la sección 9.2. Abra una hoja de trabajo en blanco. Seleccione **Stat** → **Basic Statistics** → **1-Sample Z**.

1. En el cuadro de diálogo 1-Sample Z (vea la figura A9.1), seleccione el botón de la opción **Summarized data**. Escriba **25** en el cuadro de edición Sample size. Escriba **372.5** en el cuadro de edición Mean.



**FIGURA A9.1**  
Ventana de diálogo 1-Sample Z de Minitab.

2. Escriba **15** en el cuadro de edición Standard deviation.
3. Escriba **368** en el cuadro de edición Test mean.
4. Seleccione el botón **Options**. En el cuadro de diálogo 1 Sample Z (vea la figura A9.2), en la lista desplegable Alternative, seleccione **not equal** para realizar una prueba de dos colas. (Seleccione menor que o mayor que para realizar un análisis de una cola.) Dé clic en **OK** para volver al cuadro de diálogo 1-Sample Z. Dé clic en el botón **OK**.



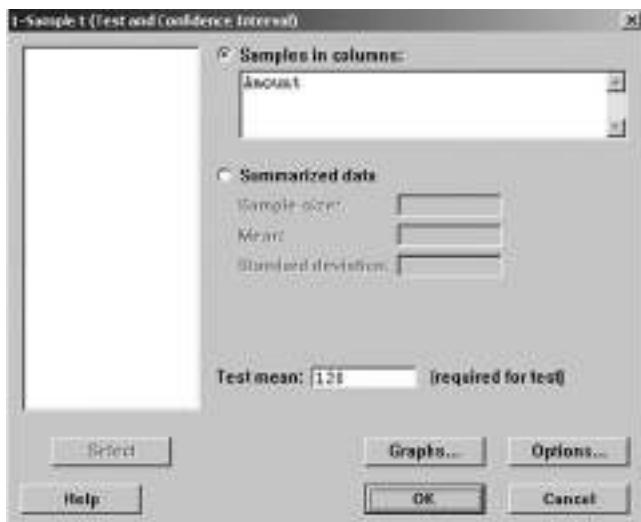
**FIGURA A9.2** Ventana de diálogo Minitab 1-Sample Z-Options.

### Uso de Minitab para la prueba $t$ de hipótesis para la media ( $\sigma$ desconocida)

Puede utilizar Minitab para una prueba de la media cuando se desconoce  $\sigma$ .

Para ilustrar la prueba de hipótesis para la media cuando no se conoce  $\sigma$ , regrese al ejemplo de las facturas de la empresa de remodelaciones Saxon analizado en la sección 9.4. Abra la hoja de trabajo INVOICES.MTW. Seleccione **Stat** → **Basic Statistics** → **1-Sample t**.

1. En la ventana de diálogo 1-Sample  $t$  (vea la figura A9.3), seleccione el botón de la opción **Sample in columns**. Escriba **C1** o **Amount** en el cuadro de edición **Sample in columns**. En el cuadro de edición **Test mean**, escriba **120**. Seleccione el botón **Options**.
2. En el menú desplegable **Alternative** del cuadro de diálogo 1-Sample  $t$  Options, seleccione **not equal** para efectuar una prueba de dos colas. (Seleccione menor que o mayor que para realizar un análisis de una cola.) Dé clic en el botón **OK** para volver a la ventana de diálogo 1-Sample  $t$  (para obtener una gráfica de caja y bigote, seleccione el botón con la opción **Graphs**. En la ventana de diálogo 1-Sample  $t$ -Graphs de Minitab, seleccione el botón con la opción



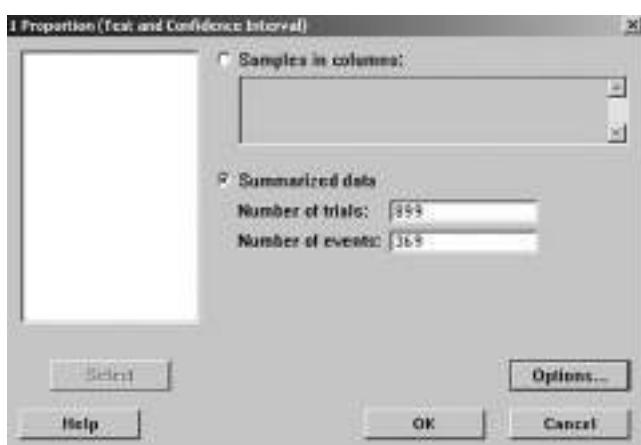
**FIGURA A9.3** Ventana de diálogo 1-Sample  $t$  de Minitab.

**Boxplot of data.** Dé clic en el botón **OK** para volver a la ventana de diálogo 1-Sample  $t$ .) Dé clic en el botón **OK**.

### Uso de Minitab para la prueba $Z$ para la proporción

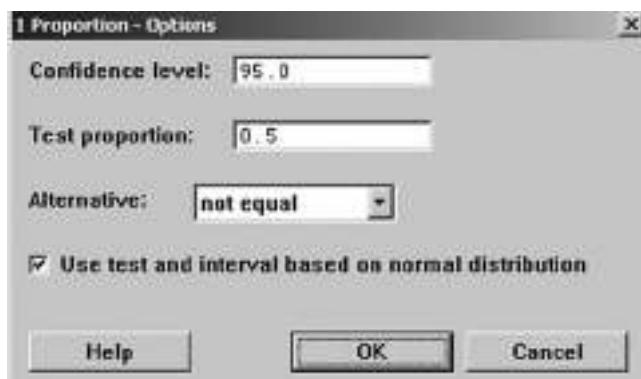
Puede utilizar Minitab para la prueba de hipótesis para la proporción. Por ejemplo, para efectuar la prueba  $Z$  para la proporción del ejemplo sobre los negocios que son propiedad de mujeres de la sección 9.5, seleccione **Stat** → **Basic Statistics** → **1 Proportion**.

1. En la ventana de diálogo 1 Proportion (vea la figura A9.4), seleccione el botón con la opción **Summarized data**. Escriba **899** en el cuadro de edición **Number of trials**. Escriba **369** en el cuadro de edición **Number of events**. Dé clic en el botón **Options**.



**FIGURA A9.4** Ventana de diálogo 1 Proportion de Minitab.

2. En la ventana de diálogo 1 Proportion-Options (vea la figura A9.5), introduzca **0.5** en el cuadro de edición **Test proportion**. Seleccione la casilla **Use test and interval based on normal distribution**. En el menú desplegable **Alternative**, seleccione **not equal** para realizar una prueba de dos colas. (Seleccione menor que o mayor que para realizar un análisis de una cola.) Dé clic en **OK** para volver a la ventana de diálogo 1 Proportion. Dé clic en el botón **OK**.



**FIGURA A9.5** Ventana de diálogo 1 Proportion-Options de Minitab.

# CAPÍTULO 10

## Pruebas de dos muestras y ANOVA de una vía

**USO DE LA ESTADÍSTICA:** Comparación de ventas de los exhibidores al final del pasillo y los exhibidores normales

### 10.1 COMPARACIÓN DE MEDIAS DE DOS POBLACIONES INDEPENDIENTES

Prueba  $Z$  para la diferencia entre dos medias  
Prueba  $t$  de varianza conjunta para diferencias entre dos medias  
Estimación del intervalo de confianza para la diferencia entre las medias de dos poblaciones independientes  
Prueba  $t$  de varianza separada para diferencia entre dos medias

### 10.2 COMPARACIÓN DE MEDIAS DE DOS POBLACIONES RELACIONADAS

Prueba  $t$  apareada  
Estimación del intervalo de confianza para la diferencia de la media

### 10.3 COMPARACIÓN DE PROPORCIONES DE DOS POBLACIONES

Prueba  $Z$  para la diferencia entre dos proporciones  
Estimación del intervalo de confianza para la diferencia entre dos proporciones

### 10.4 PRUEBA F PARA LA DIFERENCIA ENTRE DOS VARIANZAS

Encontrar los valores críticos de la cola inferior

**USO DE LA ESTADÍSTICA:** La empresa Perfect Parachute Company

### 10.5 ANOVA DE UNA VÍA

Prueba  $F$  para las diferencias entre más de dos medias  
Comparaciones múltiples: el procedimiento Tukey-Kramer  
Suposiciones ANOVA  
La prueba de la homogeneidad de la varianza de Levene

### A.10 USO DEL SOFTWARE PARA PRUEBAS DE DOS MUESTRAS Y ANOVA DE UNA VÍA

*A10.1 Excel  
A10.2 Minitab  
A10.3 (Tema del CD-ROM) SPSS*

### OBJETIVOS DE APRENDIZAJE

En este capítulo, aprenderá los procedimientos para probar hipótesis de:

- Las medias de dos poblaciones independientes
- Las medias de dos poblaciones relacionadas
- Dos proporciones
- Las varianzas de dos poblaciones independientes
- Las medias de más de dos poblaciones