

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.

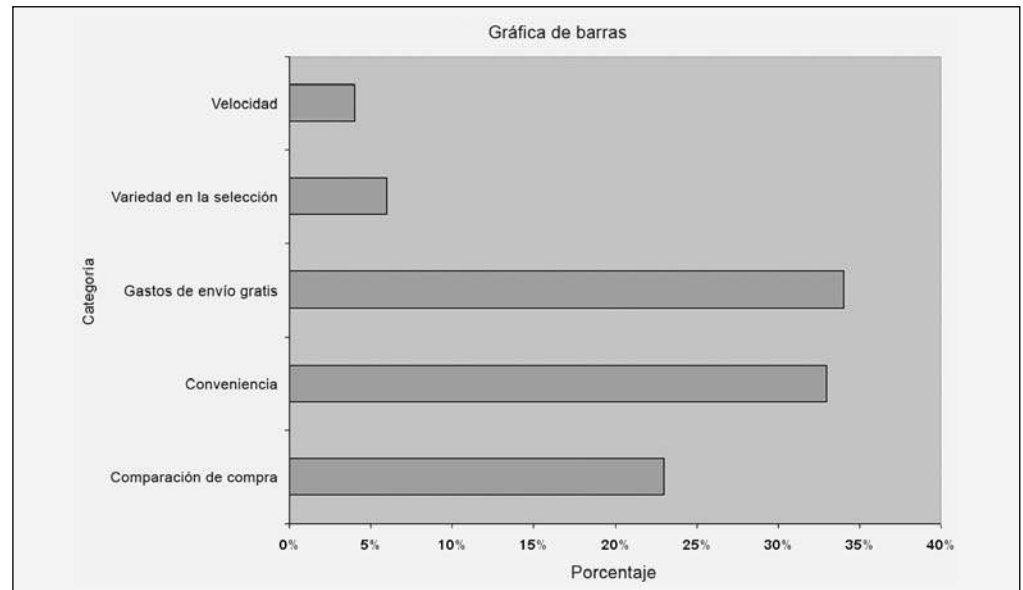
Nivel de riesgo del fondo	Número de fondos	Porcentaje de fondos
Bajo	58	47.93
Promedio	46	38.02
Alto	17	14.05
Total	121	100.00

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**

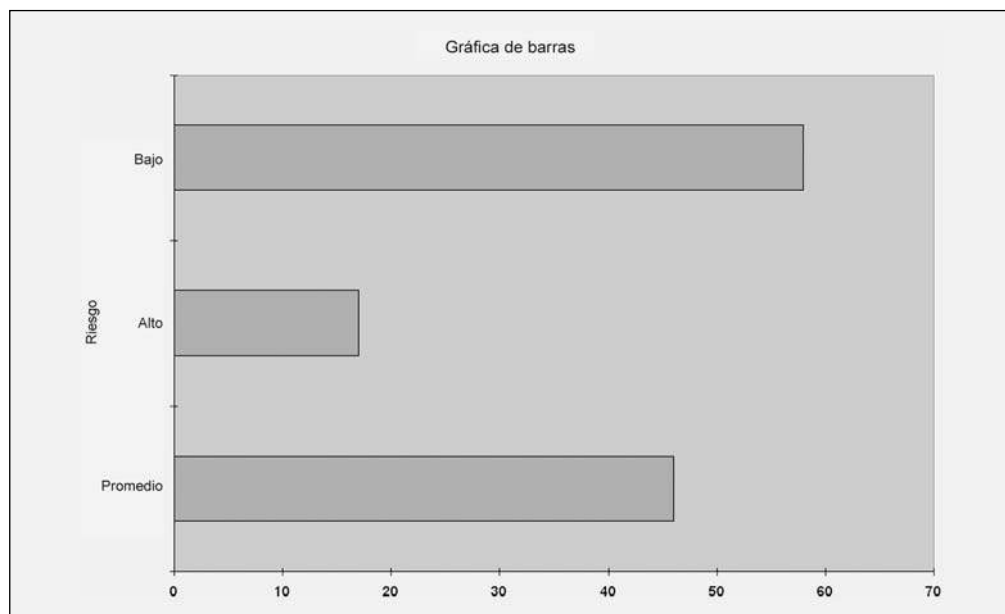
Construya una gráfica de barras para los niveles de riesgo de los fondos de inversión (con base en la información de la tabla 2.2) e interprete los resultados.

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

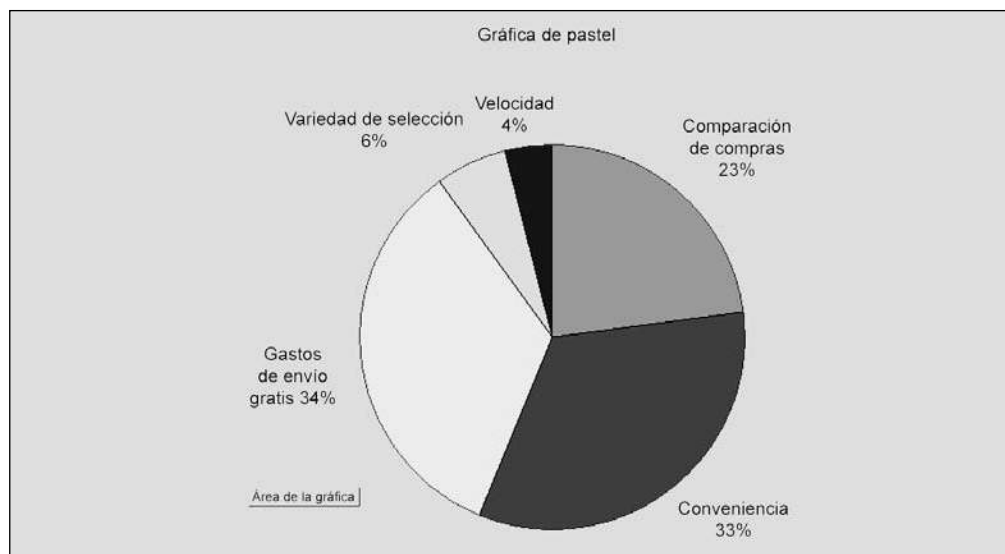
Gráfica de barras de Excel con los niveles de riesgo de los fondos de inversión.

**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° que conforman el círculo se multiplican por 0.33, de lo que resulta una rebanada del pastel que abarca 118.8° de los 360° 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.

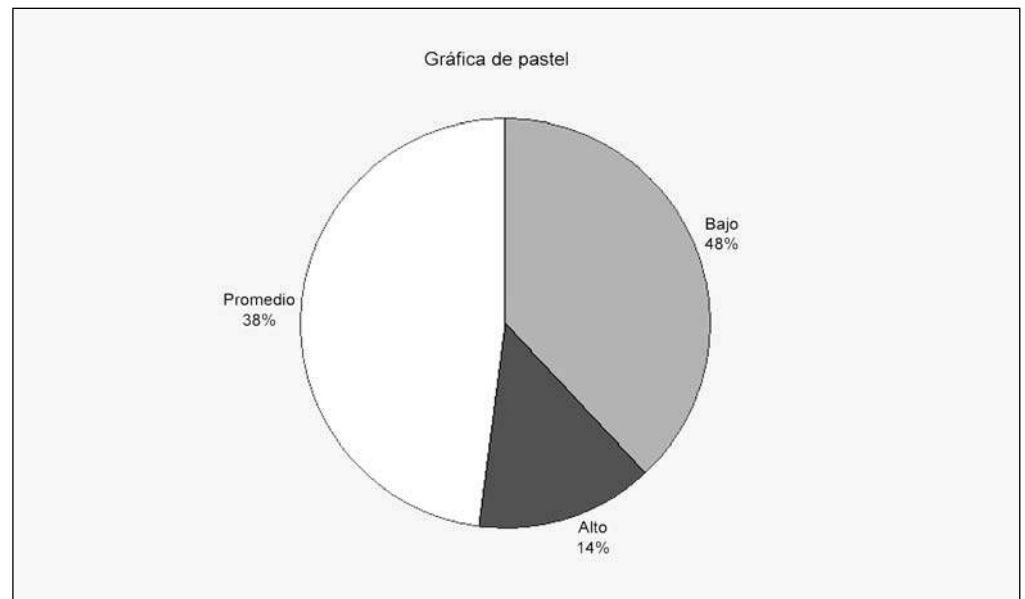


Diagrama de Pareto

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.

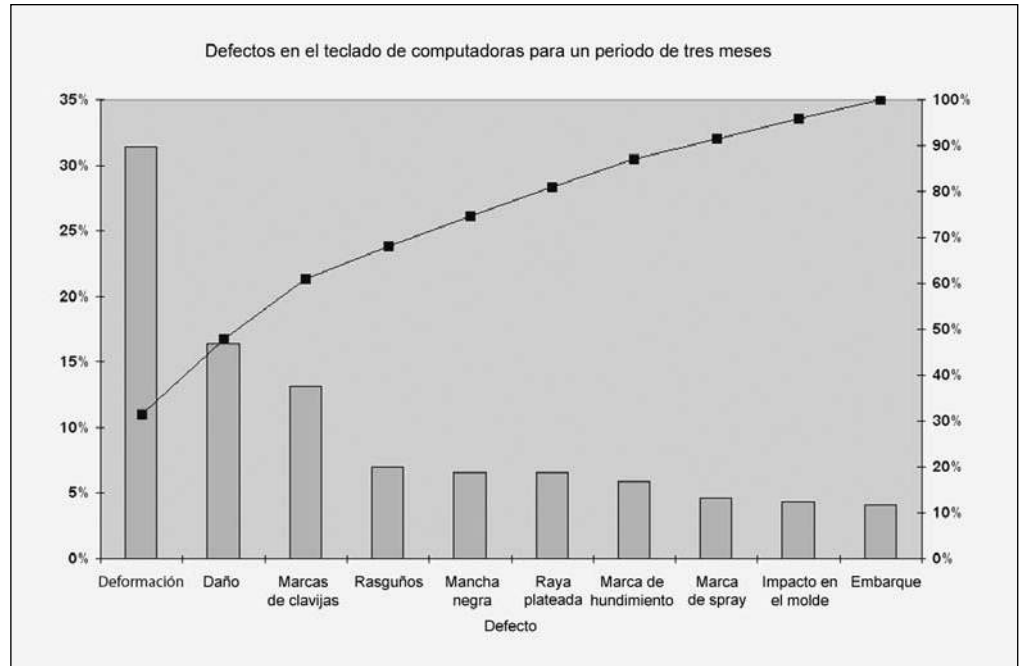
La figura 2.5 presenta las barras de forma vertical a lo largo de la línea de porcentaje acumulado.¹ La línea acumulativa está trazada en el punto medio de cada barra a una altura semejante al porcentaje acumulado. Si sigue la línea, verá que estas tres primeras categorías abarcan más del 60% de las 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. Después 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.

¹ 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

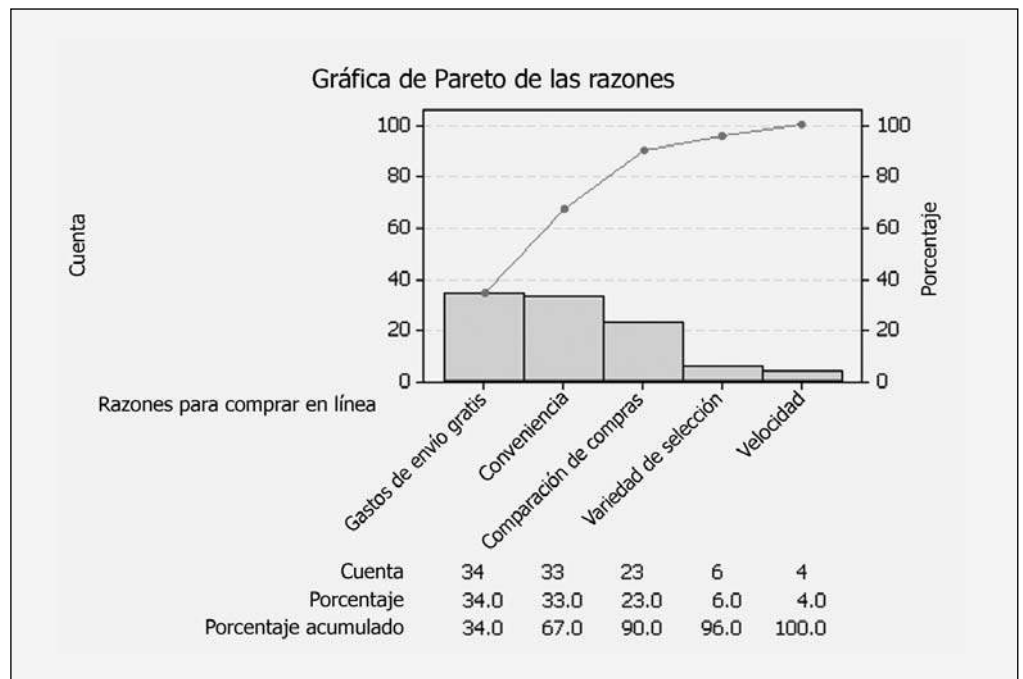
Construya un diagrama de Pareto para las razones de comprar regalos en línea (vea la tabla 2.1 en la página 22).

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

- Calcule el porcentaje de valores en cada categoría.
- Construya una gráfica de barras.
- Construya una gráfica de pastel.
- 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

- Construya una gráfica de barras.
- Construya una gráfica de pastel.
- 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

- Construya una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.
- ¿Cuál es el método gráfico que mejor refleja los datos?
- 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

- Elabore una gráfica de barras, una gráfica de pastel y un diagrama de Pareto.
- ¿Cuál método gráfico refleja mejor los datos?
- ¿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

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

- Elabore un diagrama de Pareto.
- ¿Qué porcentaje de electricidad se deriva de cualquiera de las siguientes fuentes: carbón, energía nuclear o gas natural?
- Construya una gráfica de pastel.
- ¿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 Merrill”, *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.

Ciudad									
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
Fuera de la ciudad									
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.

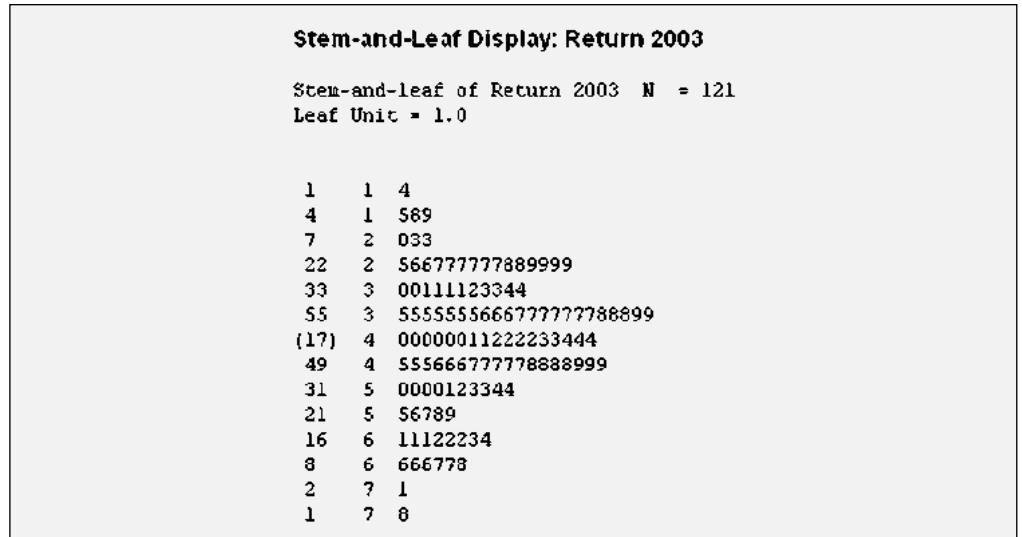
SOLUCIÓN

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
    
```

```

9 | 147
10 | 02238
11 | 125566777
12 | 223489
13 | 02
    
```

- a. Coloque los datos en un arreglo ordenado.
- b. ¿Cuál de los dos diagramas aporta más información? Explique su respuesta.
- c. ¿Cuánta gasolina (en galones) es más probable que se compre?
- d. ¿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-COSTI**

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.

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

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:

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 ciudadanos 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

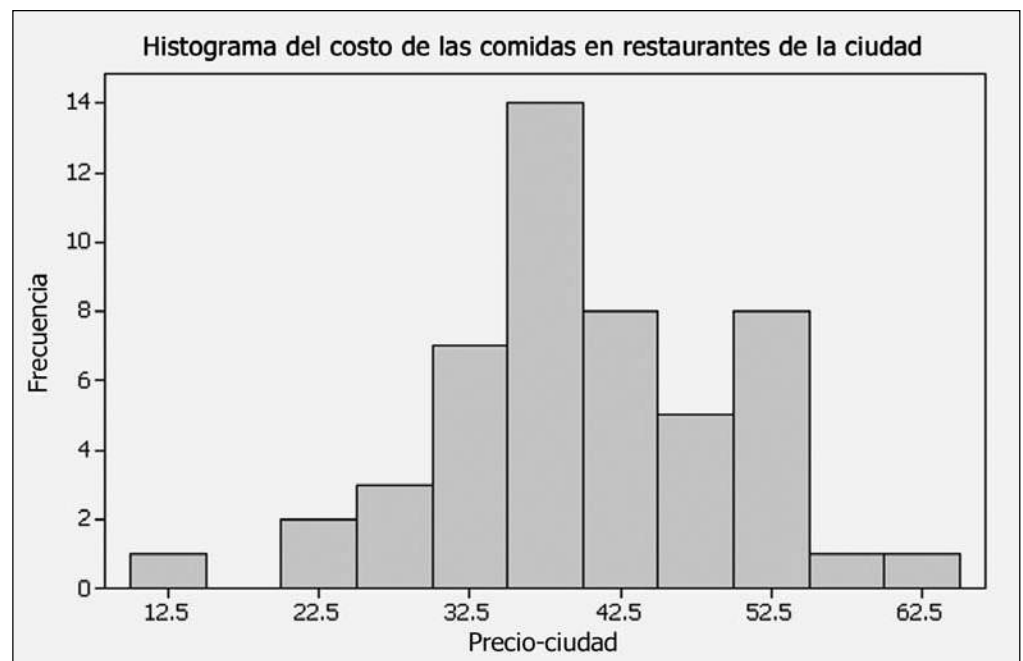
Histograma

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).

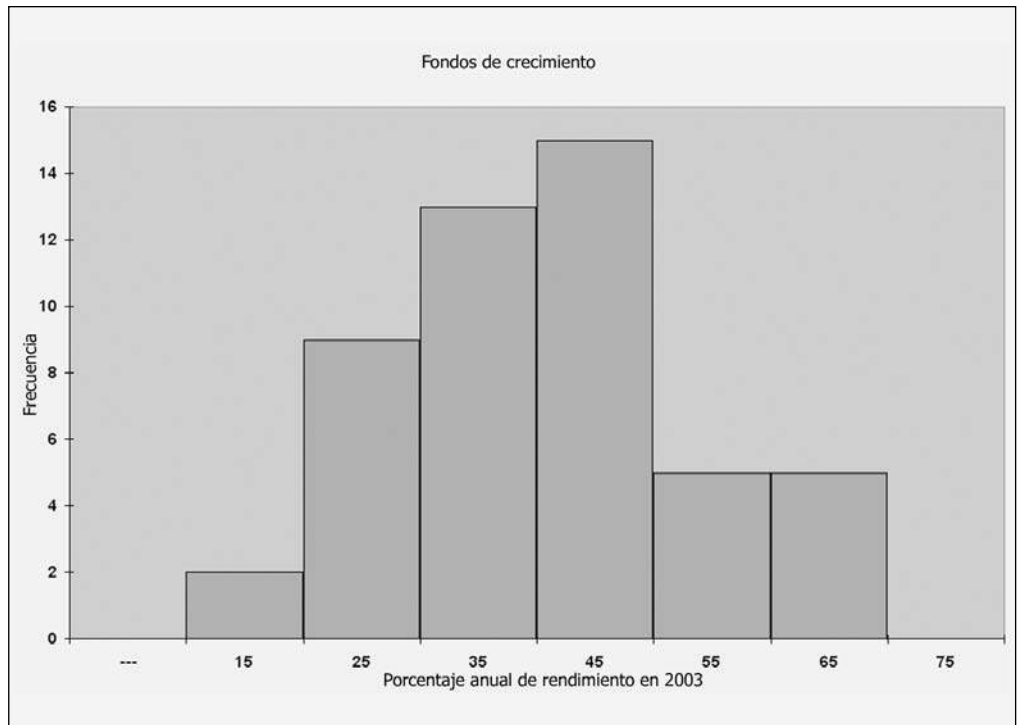


FIGURA 2.9B



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.

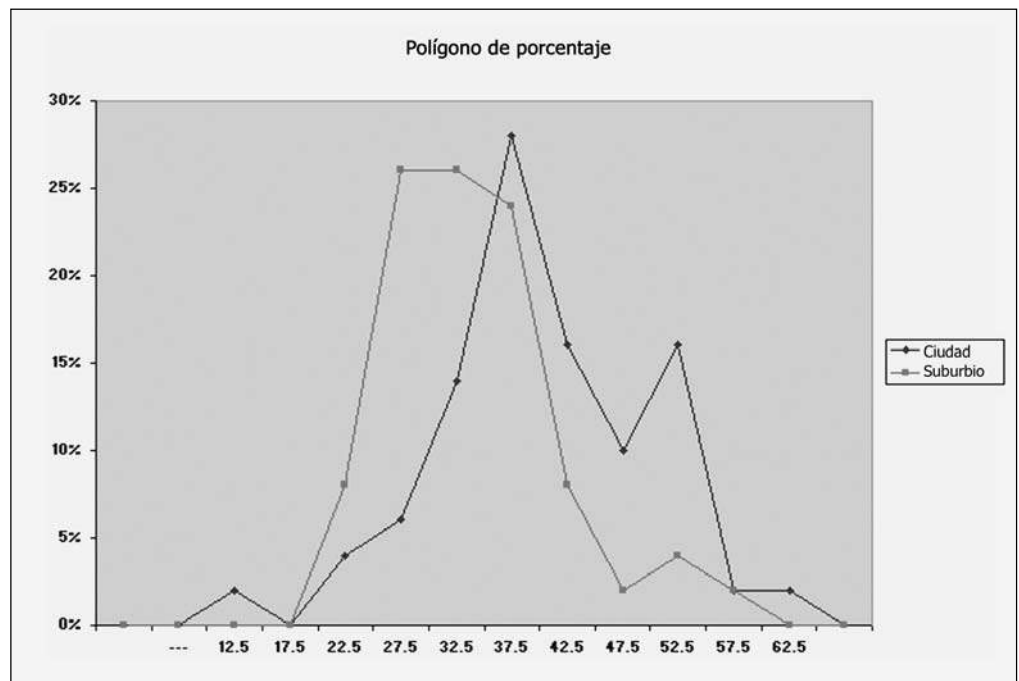
POLÍGONO DE PORCENTAJE

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

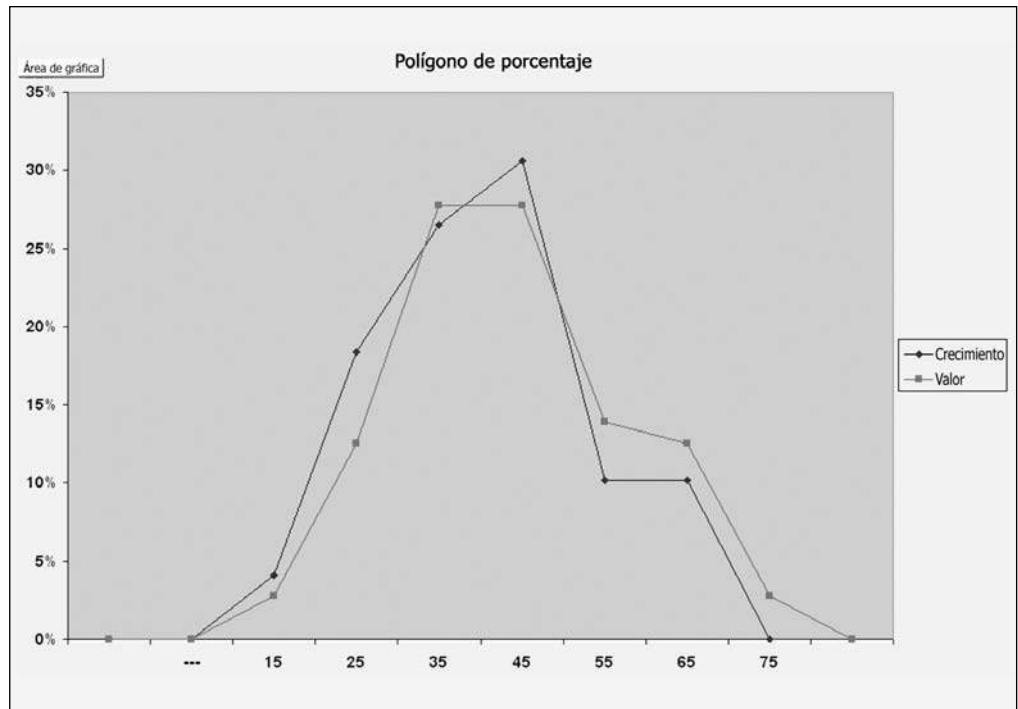
POLÍGONO DE PORCENTAJES PARA EL RENDIMIENTO EN 2003 DE LOS FONDOS DE INVERSIÓN DE CRECIMIENTO Y DE VALOR

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.

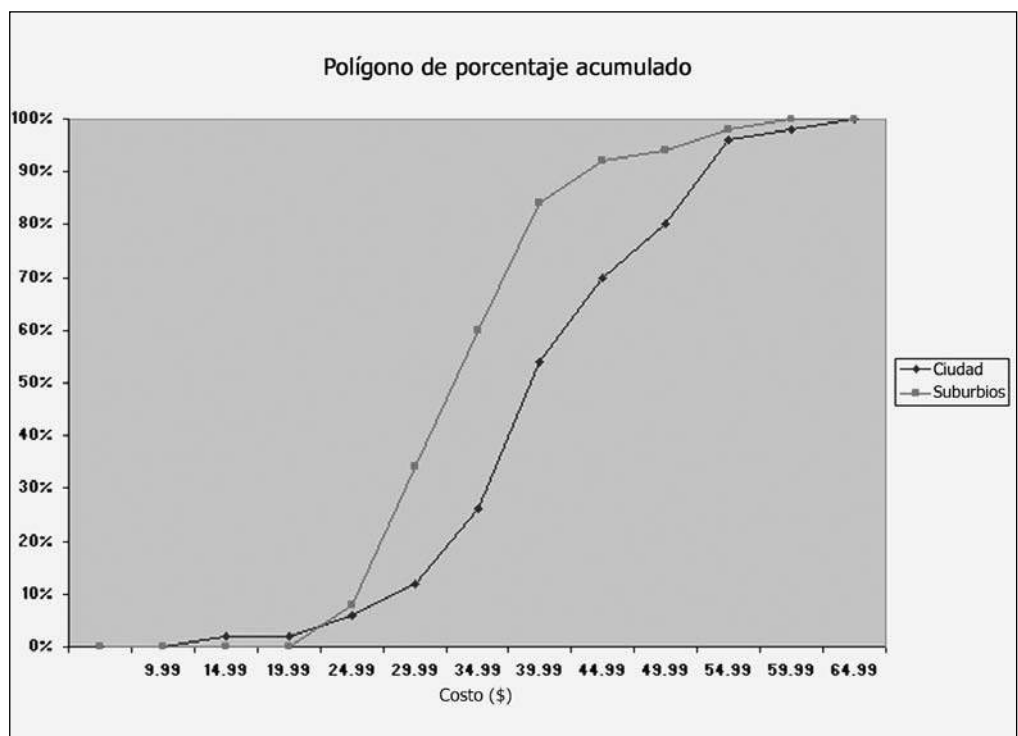


Polígono de porcentaje acumulado (ojiva)

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 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

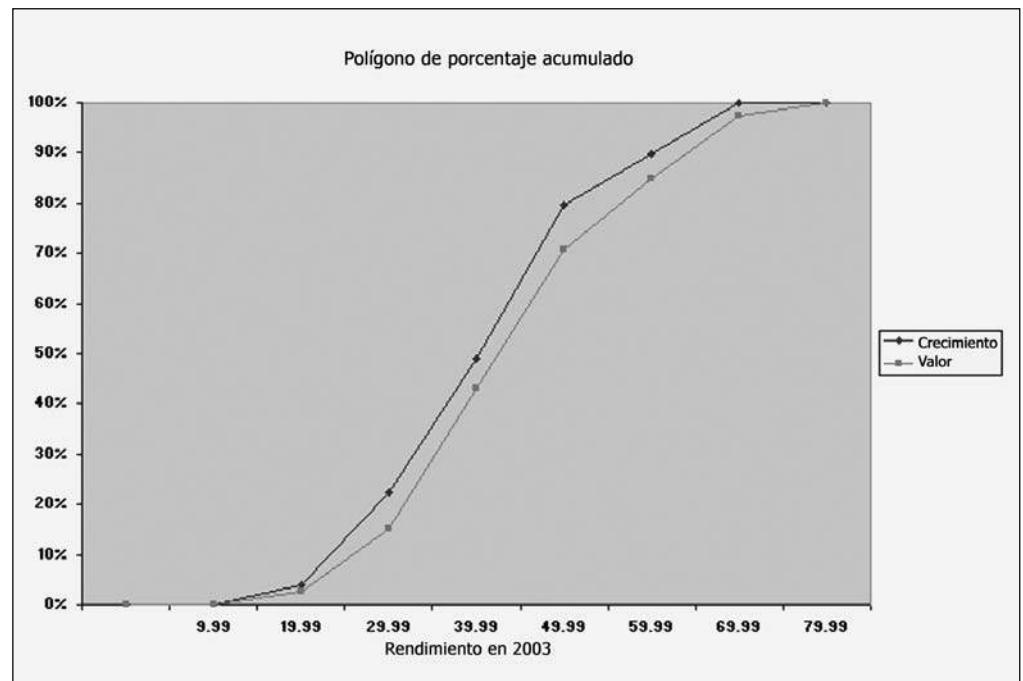
En el escenario “Uso de la estadística”, se le pide que compare el rendimiento de los fondos de inversión de crecimiento y de valor. **MUTUALFUNDS2004** Elabore polígonos de porcentaje acumulado para los fondos de crecimiento y para los fondos de valor.

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 sucesiva-

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:

- a. ¿Qué porcentaje de solicitantes calificó por debajo de 500?
- b. ¿Qué porcentaje de solicitantes calificó entre 500 y 549?
- c. ¿Qué porcentaje de solicitantes calificó por debajo de 550?
- d. ¿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 de PH Grade **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

- a. 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.
- b. Trace un histograma y un polígono de porcentaje.
- c. Elabore una distribución de porcentaje acumulado y trace una ojiva (polígono de porcentaje acumulado).
- d. ¿Alrededor de cuál cantidad parece concentrarse el costo mensual de la energía eléctrica?

AUTO Examen **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 ± 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

- a. Realice una distribución de frecuencias y una distribución de porcentaje.
- b. Trace un histograma y un polígono de porcentaje.
- c. Trace un polígono de porcentaje acumulado.
- d. ¿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							

- a. Realice una distribución de frecuencia y una distribución de porcentajes.
- b. Trace un histograma y un polígono de porcentajes.
- c. Trace un polígono de porcentajes acumulados.
- d. ¿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		

- a. Construya una distribución de frecuencias y una distribución de porcentajes.
- b. Trace un histograma y un polígono de porcentajes.
- c. Trace un polígono de porcentaje acumulado.
- d. ¿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

- a. Realice una distribución de frecuencias y una distribución de porcentajes para cada fabricante usando el siguiente intervalo de clase para cada distribución:
 - (1) Fabricante A: 650 pero menos de 750, 750 pero menos de 850, y así sucesivamente.
 - (2) Fabricante B: 750 pero menos de 850, 850 pero menos de 950, y así sucesivamente.
- b. Trace los histogramas de porcentaje en gráficas separadas y trace los polígonos de porcentaje en una gráfica.
- c. Elabore las distribuciones de porcentaje acumulado y trace las ojivas en una gráfica.
- d. ¿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						

- a. Construya una distribución de frecuencias y una distribución de porcentajes.
- b. Trace un histograma y un polígono de porcentajes.
- c. Realice una distribución de porcentaje acumulado y trace un polígono de porcentaje acumulado.
- d. 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.

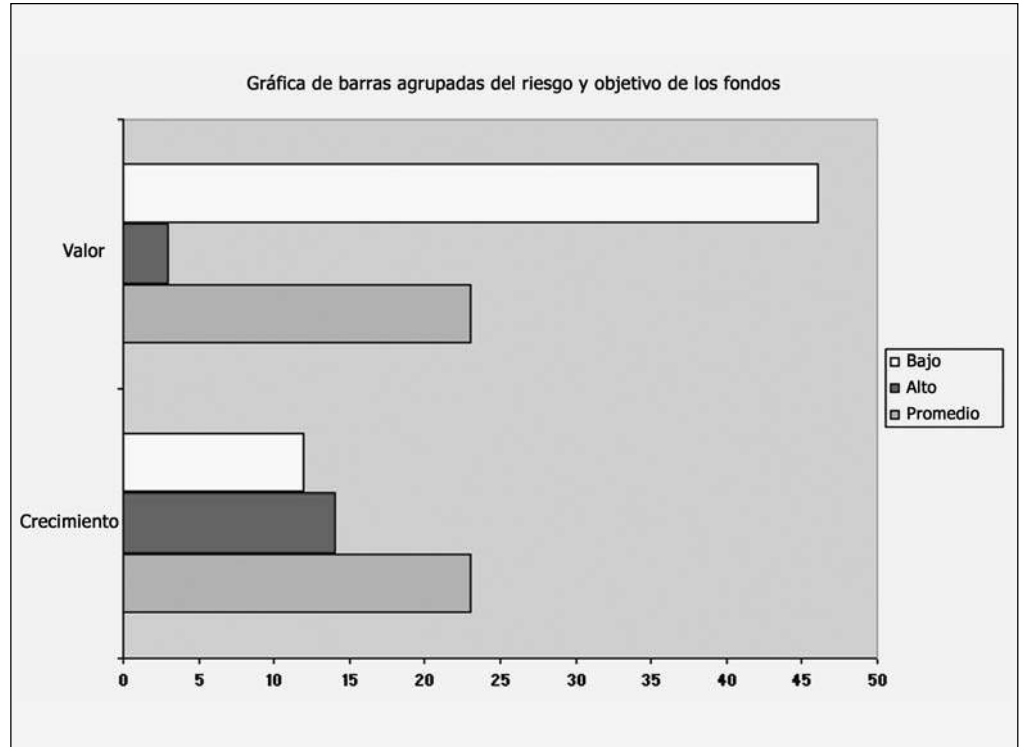
OBJETIVO	NIVEL DE RIESGO			Total
	Alto	Promedio	Bajo	
Crecimiento	82.35	50.00	20.69	40.50
Valor	17.65	50.00	79.31	59.50
Total	100.00	100.00	100.00	100.00

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
<i>A</i>	20	40	40	100
<i>B</i>	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
Totales	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 llegará 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.

c. ¿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. Lundegaard 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

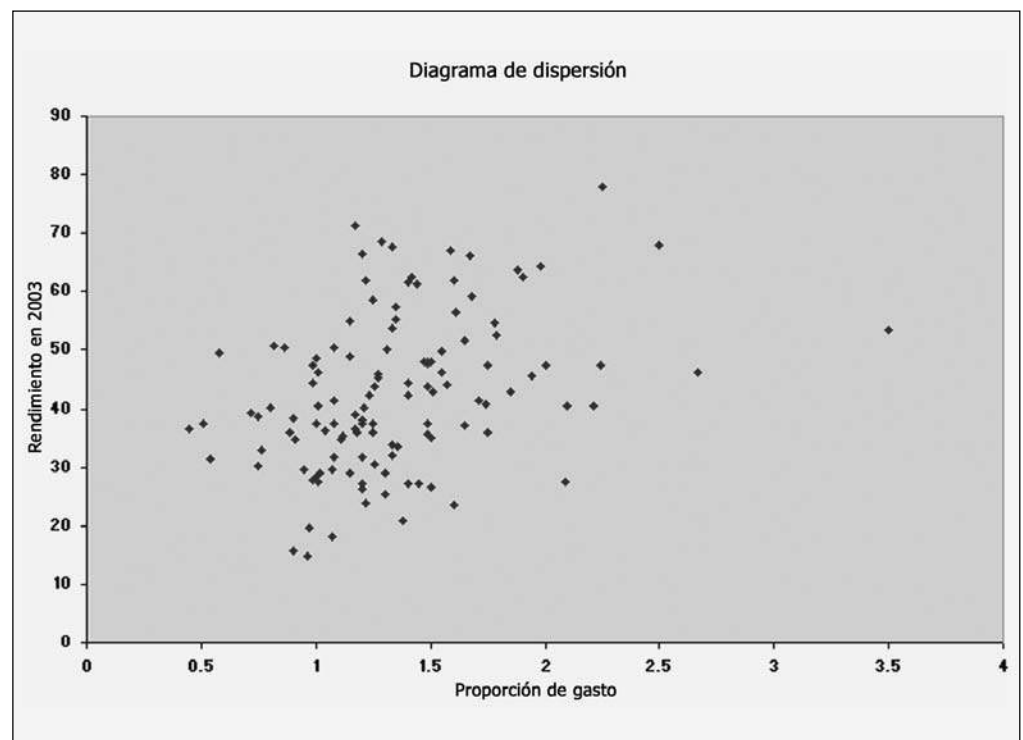
Diagrama de dispersión

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.

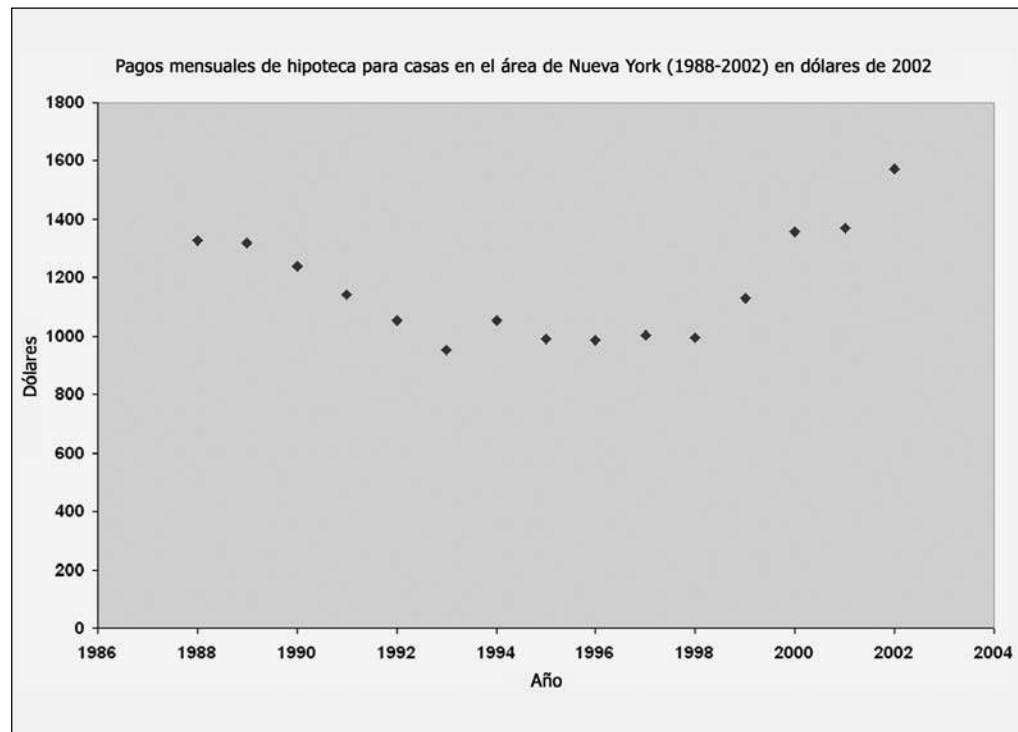
Series de tiempo

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.

X	7	5	8	3	6	10	12	4	9	15	18
Y	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).

Año	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Ventas	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
Kenmore7118	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

- 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**

Tasa de desempleo en E.U. ajustada por temporada (en porcentaje)						
Mes	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-TV**

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?

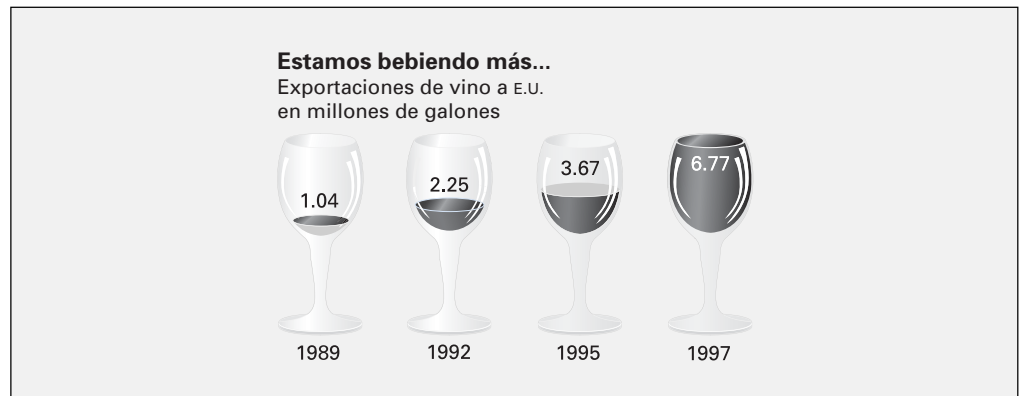
2.6 USO INADECUADO DE GRÁFICAS Y CONSIDERACIONES ÉTICAS

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.

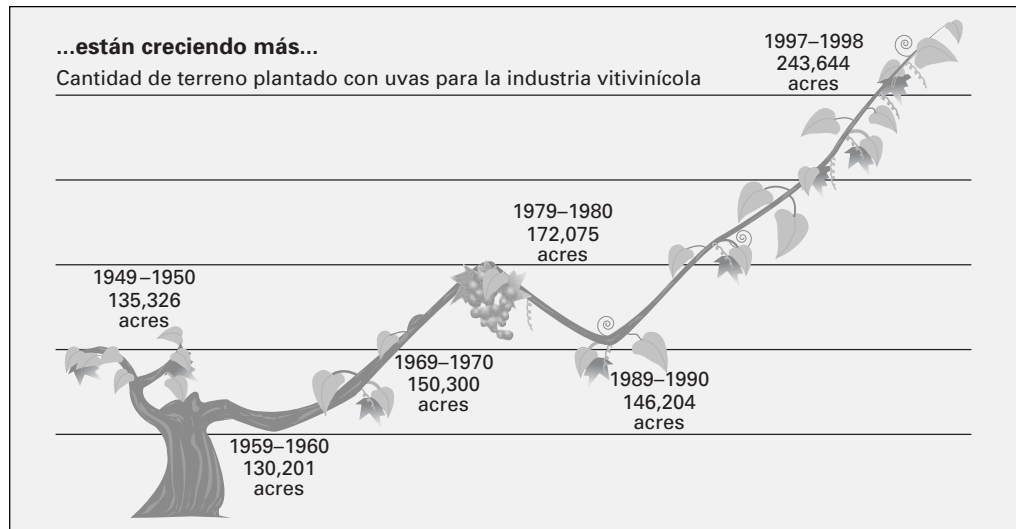
Además del tipo de distorsión creada por los iconos de la copa de vino en la gráfica de la revista *Time* que muestra la figura 2.17, el uso impropio de los ejes vertical y horizontal lleva a distorsiones. La figura 2.18 en la página 52 presenta otra gráfica usada en el mismo artículo de *Time*.

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 ocupa 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

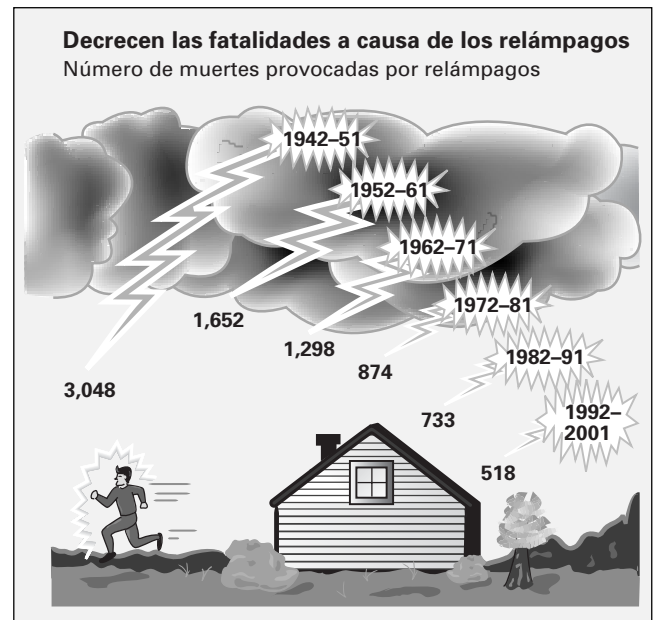
Aplicación de conceptos

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.

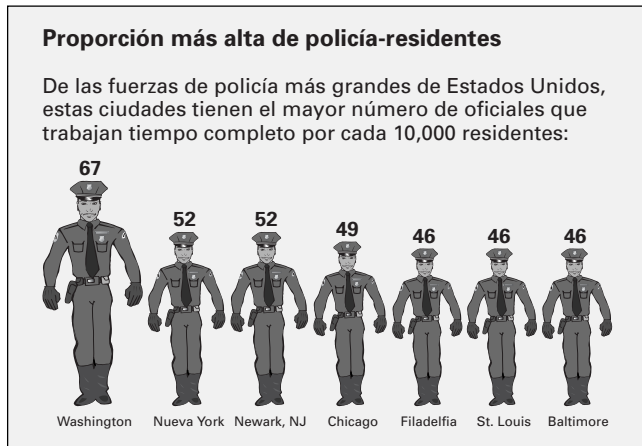
2.49 La siguiente muestra visual contiene una gráfica sobreadornada que apareció en el *USA Today* y que se refiere a las muertes a causa de los relámpagos en Estados Unidos.



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.

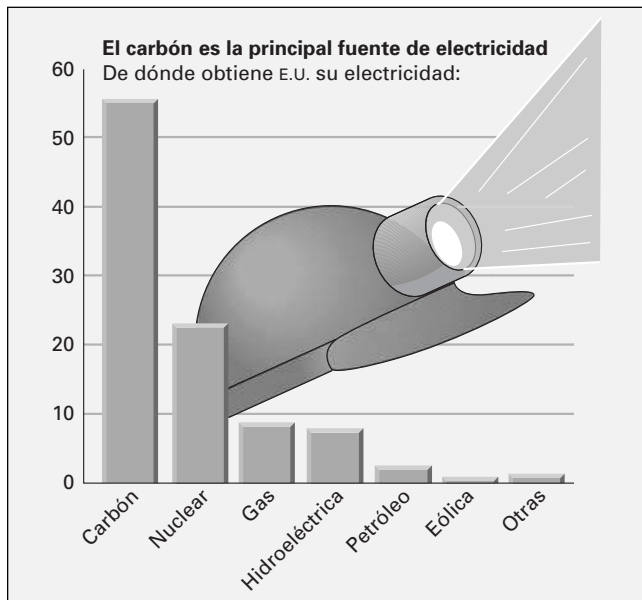
2.50 La siguiente muestra visual se refiere al tamaño relativo de los departamentos de policía en las ciudades más grandes de Estados Unidos, que apareció en el *USA Today*:



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.

- 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.

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-

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	Catagóricos
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 (secciones 2.2 y 2.3)	Tabla de resumen, gráfica de barras, gráfica de pastel, diagrama de Pareto (sección 2.1)
Presentar gráficamente la relación entre dos variables	Diagrama de dispersión, Series de tiempo (sección 2.5)	Tabla de contingencia, gráfica de barras agrupadas (sección 2.4)

CONCEPTOS CLAVE

Agrupaciones o clases 33	Distribución de porcentaje acumulado 35	Diagrama de series de tiempo 48
Amplitud de un intervalo de clase 33	Distribución de porcentajes 34	Polígono de porcentaje 39
Arreglo ordenado 30	Gráfica de barras 23	Polígono de porcentaje acumulado 40
Basura gráfica 52	Gráfica de barras agrupadas 45	Principio de Pareto 25
Celdas 43	Gráfica de pastel 24	Punto medio de la clase 33
Diagrama de tallo y hoja 30	Histograma 37	Rango 33
Diagrama de dispersión 47	Límites de clase 33	Tabla de datos bivariados 43
Diagrama de Pareto 25	Ojiva (polígono de porcentaje acumulado) 40	Tabla de contingencia 43
Distribución de frecuencia relativa 34		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 catagóricos 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	Reservas
<i>Norteamérica</i>	54.8	
México		28.3
E.U.		21.8
Canadá		4.7
<i>Centro y Sudamérica</i>	95.2	
Venezuela		76.9
Brasil		8.1
Otros en Centro y Sudamérica		10.2
<i>Europa Occidental</i>	17.2	
Noruega		9.5
Inglaterra		5.0
Otros en Europa Occidental		2.7
<i>África</i>	74.9	
Libia		29.5
Nigeria		22.5
Argelia		9.2
Angola		5.4
Otros en África		8.3

Reservas convencionales probadas (en miles de millones de barriles)

Región y país	Reservas
<i>Medio Oriente</i>	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
<i>Lejano Oriente y Oceanía</i>	44.0
China	24.0
Indonesia	5.0
India	4.7
Otros en Lejano Oriente y Oceanía	10.3
<i>Europa Oriental y ex URSS</i>	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:

ORDENÓ POSTRE	GÉNERO		Total
	Masculino	Femenino	
Sí	96	40	136
No	224	240	464
Total	320	280	600

ORDENÓ POSTRE	PLATILLO DE RES		Total
	Sí	No	
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.

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

- Elabore una distribución de frecuencia y una distribución de porcentaje.
- Trace un histograma y un polígono de porcentaje.
- Construya una distribución de porcentaje acumulado y trace una ojiva (polígono de porcentaje acumulado).
- 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.

- Elabore distribuciones de frecuencias y distribuciones de porcentaje para grasa, costo y calorías.
- 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 un 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 (carne roja fresca, 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, precio 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*

- 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.