

# CAPÍTULO 2



## Estadística descriptiva: presentaciones tabulares y gráficas

---

### CONTENIDO

ESTADÍSTICA EN LA PRÁCTICA:  
COLGATE-PALMOLIVE COMPANYY

#### 2.1 RESUMEN DE DATOS CUALITATIVOS

Distribución de frecuencia  
Distribuciones de frecuencia  
relativa y frecuencia porcentual  
Gráficas de barras y circulares

#### 2.2 RESUMEN DE DATOS CUANTITATIVOS

Distribución de frecuencia  
Distribuciones de frecuencia  
relativa y frecuencia porcentual  
Diagramas de puntos

Histograma

Distribuciones acumuladas  
Ojiva

#### 2.3 ANÁLISIS DE DATOS EXPLORATORIOS: EL DIAGRAMA DE TALLO Y HOJA

#### 2.4 TABULACIONES CRUZADAS Y DIAGRAMAS DE DISPERSIÓN

Tabulación cruzada  
La paradoja de Simpson  
Diagrama de dispersión y línea  
de tendencia

## ESTADÍSTICA *en* LA PRÁCTICA

### COLGATE-PALMOLIVE COMPANY\* NEW YORK, NEW YORK

Colgate-Palmolive inició como una tienda pequeña de jabones y velas en la ciudad de Nueva York en 1806. Hoy la empresa emplea a más de 40 000 personas que trabajan en más de 200 países y territorios de todo el mundo. Aunque es más conocida por sus marcas Colgate, Palmolive, Ajax y Fab, la empresa también comercializa los productos de Mennen, Hill's Science Diet y Hill's Prescription Diet.

Colgate-Palmolive aplica la estadística en su programa de aseguramiento de la calidad para los productos de detergente para ropa. Una preocupación radica en que el cliente quede satisfecho con la cantidad de contenido que se incluye en cada paquete. Los empaques de cada categoría se llenan con la misma cantidad de detergente en peso, pero el volumen varía dependiendo de la densidad del polvo. Por ejemplo, si ésta es alta, se necesita un volumen menor del producto para alcanzar el peso especificado en el empaque. Como resultado, cuando el consumidor lo abre, le parece que le falta detergente.

Para controlar el problema de la alta densidad del detergente en polvo, se han establecido límites aceptables para este nivel. De manera periódica se toman muestras estadísticas del producto y se mide la densidad de cada muestra. Luego se proporcionan resúmenes de los datos a los operarios para que emprendan acciones correctivas en caso necesario con el fin de mantener la densidad dentro de las especificaciones de calidad deseadas.

En la tabla y figura adjuntas se presenta una distribución de frecuencia y un histograma de las densidades de 150 muestras tomadas durante un periodo de una semana. Los niveles de densidad superiores a 0.40 son inaceptablemente altos. La distribución de frecuencia y el histograma indican que la operación cumple con las especificaciones de calidad, ya que todas las densidades son menores o iguales a 0.40. Cuando los gerentes leen estos resúmenes estadísticos quedan complacidos con la calidad del proceso de producción del detergente.

En este capítulo usted aprenderá acerca de los métodos tabulares y gráficos de la estadística descriptiva, por ejemplo las distribuciones de frecuencia, las gráficas de barras, los histogramas, los diagramas de tallo y hoja y las tabulaciones cruzadas, entre otros. El objetivo de estos métodos es resumir los datos de modo que sea fácil entenderlos e interpretarlos.



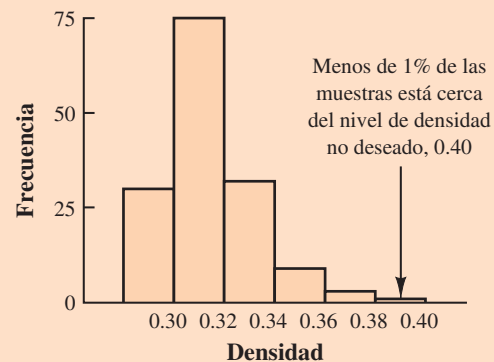
Los resúmenes gráficos ayudan a seguir la pista de la demanda de los productos de Colgate-Palmolive.

© Victor Fisher/ Bloomberg News/Landov.

**Distribución de frecuencia  
de los datos de densidad**

Densidad	Frecuencia
0.29–0.30	30
0.31–0.32	75
0.33–0.34	32
0.35–0.36	9
0.37–0.38	3
0.39–0.40	1
Total	150

**Histograma de los datos de densidad**



\* Los autores agradecen a William R. Fowle, gerente de Aseguramiento de la Calidad de Colgate-Palmolive, por proporcionar este artículo.

Como se indicó en el capítulo 1, los datos se clasifican en cualitativos o cuantitativos. Los **datos cualitativos** utilizan etiquetas o nombres para identificar las categorías de elementos similares. Los **datos cuantitativos** son valores numéricos que indican cuánto o cuántos.

Este capítulo presenta los métodos tabulares y gráficos de uso común para resumir datos cualitativos y cuantitativos. Los resúmenes tabulares y gráficos de los datos pueden encontrarse en informes anuales, artículos periodísticos y estudios de investigación. Todos estamos expuestos a este tipo de presentaciones. Por tanto, es importante entender cómo se elaboran y cómo deben interpretarse. Comencemos con los métodos tabulares y gráficos para resumir los datos que se refieren a una sola variable. La última sección presenta los métodos para resumir datos cuando lo que interesa es la relación entre dos variables.

El software moderno para estadística cuenta con numerosas funciones para resumir datos y elaborar presentaciones gráficas. Minitab y Excel son dos paquetes que se utilizan mucho. En los apéndices del capítulo se mencionan algunas de sus funciones.

## 2.1

## Resumen de datos cualitativos

### Distribución de frecuencia

Con el fin de explicar cómo se usan los métodos tabulares y gráficos para resumir datos cualitativos, comenzaremos con la definición de **distribución de frecuencia**.

#### DISTRIBUCIÓN DE FRECUENCIA

Una distribución de frecuencia es un resumen tabular de datos que muestra el número (frecuencia) de elementos en cada una de varias clases que no se superponen.

Con el ejemplo siguiente se explica la elaboración e interpretación de una distribución de frecuencia para datos cualitativos. Coke Classic, Diet Coke, Dr. Pepper, Pepsi y Sprite son cinco bebidas refrescantes conocidas. Suponga que los datos de la tabla 2.1 presentan la bebida refrescante seleccionada en una muestra de 50 bebidas adquiridas.

**TABLA 2.1** Datos de una muestra de 50 bebidas refrescantes adquiridas

Coke Classic	Sprite	Pepsi
Diet Coke	Coke Classic	Coke Classic
Pepsi	Diet Coke	Coke Classic
Diet Coke	Coke Classic	Coke Classic
Coke Classic	Diet Coke	Pepsi
Coke Classic	Coke Classic	Dr. Pepper
Dr. Pepper	Sprite	Coke Classic
Diet Coke	Pepsi	Diet Coke
Pepsi	Coke Classic	Pepsi
Pepsi	Coke Classic	Pepsi
Coke Classic	Coke Classic	Pepsi
Dr. Pepper	Pepsi	Pepsi
Sprite	Coke Classic	Coke Classic
Coke Classic	Sprite	Dr. Pepper
Diet Coke	Dr. Pepper	Pepsi
Coke Classic	Pepsi	Sprite
Coke Classic	Diet Coke	

**TABLA 2.2**

Distribución de frecuencia de la compra de bebidas refrescantes

Bebida refrescante	Frecuencia
Coke Classic	19
Diet Coke	8
Dr. Pepper	5
Pepsi	13
Sprite	5
Total	50

Para elaborar una distribución de frecuencia con estos datos, se cuenta el número de veces que cada bebida refrescante aparece en la tabla 2.1. La Coke Classic se registra 19 veces, la Diet Coke 8, Dr. Pepper 5, Pepsi 13 y Sprite 5. Estos conteos se resumen en la tabla 2.2.

Esta distribución de frecuencia proporciona un resumen de cómo se distribuyen las 50 bebidas refrescantes adquiridas entre las cinco marcas. El resumen ofrece una visión más clara que los datos originales mostrados en la tabla 2.1. Al observar la distribución de frecuencia, Coke Classic destaca como la bebida refrescante más vendida, Pepsi como la segunda, Diet Coke la tercera, y Sprite y Dr. Pepper empatan en el cuarto lugar. La distribución de frecuencia resume la información acerca de la popularidad de las cinco bebidas.

## Distribuciones de frecuencia relativa y frecuencia porcentual

Una distribución de frecuencia muestra el número (la frecuencia) de elementos en cada una de varias clases que no se superponen. Sin embargo, lo que interesa a menudo es la proporción, o el porcentaje, de elementos en cada clase. La *frecuencia relativa* de una clase es igual a la fracción o proporción de elementos que pertenecen a cada clase. Para un conjunto de datos con  $n$  observaciones, la frecuencia relativa de cada clase se determina como sigue.

### FRECUENCIA RELATIVA

$$\text{Frecuencia relativa de una clase} = \frac{\text{frecuencia de la clase}}{n} \quad (2.1)$$

La *frecuencia porcentual* de una clase es la frecuencia relativa multiplicada por 100.

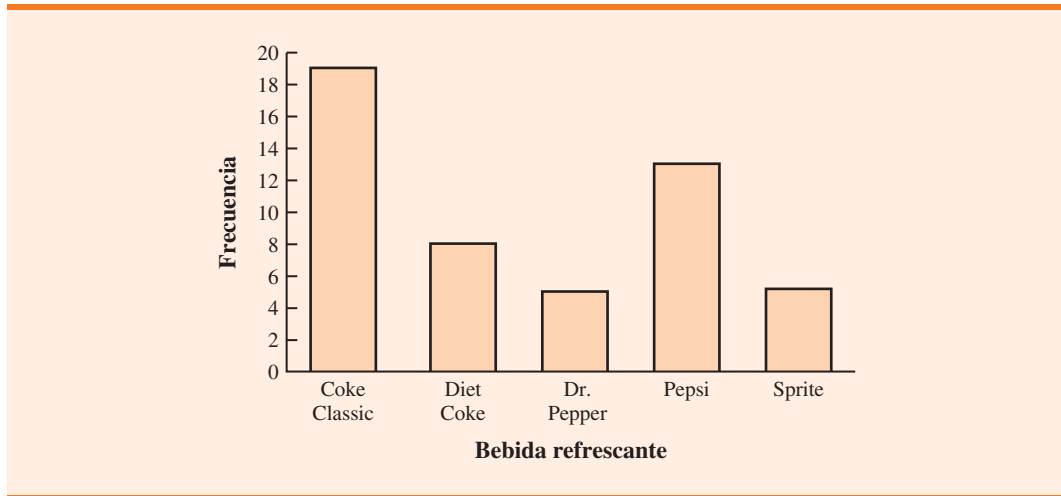
Una **distribución de frecuencia relativa** proporciona un resumen tabular de los datos que indica la frecuencia relativa de cada clase. Una **distribución de frecuencia porcentual** resume la frecuencia porcentual de los datos para cada clase. La tabla 2.3 muestra las distribuciones de frecuencia relativa y frecuencia porcentual para los datos de las bebidas refrescantes. En ella se aprecia que la frecuencia relativa de la Coke Classic es  $19/50 = 0.38$ , la frecuencia relativa de la Diet Coke es  $8/50 = 0.16$ , etc. En la distribución de frecuencia porcentual se aprecia que 38% de las bebidas refrescantes adquiridas fue de Coke Classic, 16% de Diet Coke, etc. También se observa que  $38\% + 26\% + 16\% = 80\%$  de las bebidas refrescantes compradas fue de las tres marcas principales de la muestra.

## Gráficas de barras y circulares

Una **gráfica de barras** es un dispositivo gráfico que se usa para representar los datos cualitativos resumidos en una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. En un eje de la gráfica (por lo general el horizontal) se especifican las etiquetas utilizadas para las clases (categorías). En el otro eje de la gráfica (por lo general el vertical)

**TABLA 2.3** Distribuciones de frecuencia relativa y frecuencia porcentual de la compra de bebidas refrescantes

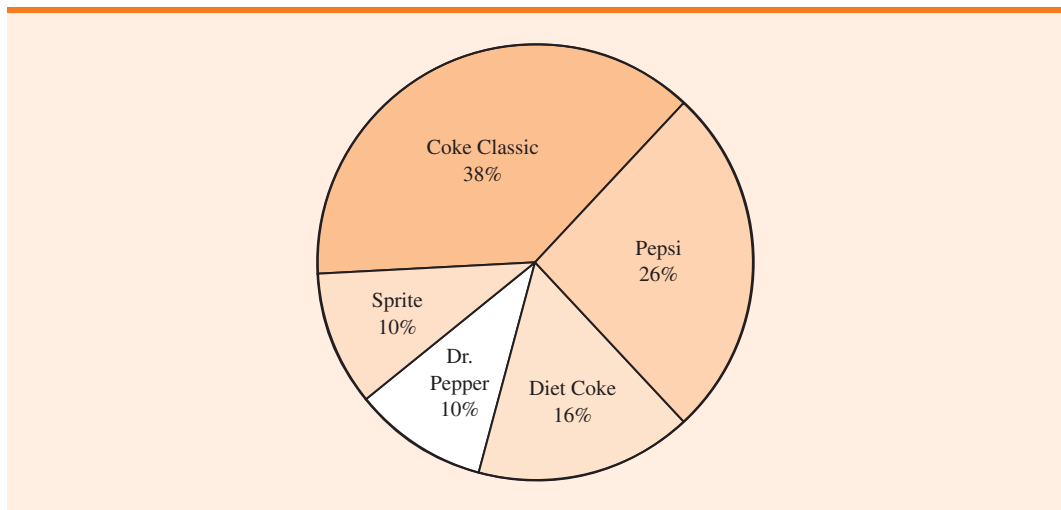
Bebida refrescante	Frecuencia relativa	Frecuencia porcentual
Coke Classic	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	0.10	10
Total	1.00	100

**FIGURA 2.1** Gráfica de barras de la compra de bebidas refrescantes

Las gráficas de barras se usan en las aplicaciones de control de calidad para identificar las principales causas de los problemas. Cuando las barras se acomodan en orden descendente de altura y de izquierda a derecha, colocando primero la causa que ocurre con más frecuencia, la gráfica de barras se llama diagrama de Pareto. Recibe este nombre en honor de su fundador, Wilfredo Pareto, un economista italiano.

se coloca una escala de frecuencia, de frecuencia relativa o de frecuencia porcentual. Luego se traza una barra de ancho fijo arriba de cada etiqueta de clase y se extiende su longitud hasta llegar a la frecuencia, la frecuencia relativa o la frecuencia porcentual de la clase. Para los datos cualitativos, las barras deben estar a cierta distancia para recalcar el hecho de que cada clase está separada. La figura 2.1 muestra una gráfica de barras de la distribución de frecuencia de las 50 bebidas refrescantes adquiridas. Note cómo la presentación muestra que la Coke Classic, la Pepsi y la Diet Coke son las marcas preferidas.

La **gráfica circular o de pastel** es otro dispositivo gráfico que presenta las distribuciones de frecuencia relativa y de frecuencia porcentual para datos cualitativos. Para elaborarla, primero se traza un círculo que represente todos los datos. Luego se usan las frecuencias relativas para subdividir el círculo en sectores, o partes, que corresponden a la frecuencia relativa de cada clase. Por ejemplo, como un círculo contiene 360 grados y la Coke Classic muestra una frecuencia relativa de 0.38, el sector de la gráfica circular etiquetado como Coke Classic mide  $0.38(360) = 136.8$  grados. El sector etiquetado como Diet Coke mide  $0.16(360) = 57.6$  grados. Cálculos parecidos para las demás clases producen la gráfica circular de la figura 2.2. Los

**FIGURA 2.2** Gráfica circular de la compra de bebidas refrescantes

valores numéricos mostrados para cada sector pueden ser frecuencias, frecuencias relativas o frecuencias porcentuales.

## NOTAS Y COMENTARIOS

1. A menudo, el número de clases en una distribución de frecuencia es igual al número de categorías encontradas en los datos, como ocurre con los de la compra de bebidas refrescantes en esta sección. Los datos se refieren sólo a cinco marcas, y para cada una se definió una clase de distribución de frecuencia separada. Si los datos hicieran referencia a todas las bebidas, se requerirían muchas categorías, la mayoría de las cuales tendría un número pequeño de bebidas refrescantes adquiridas. La mayoría de los expertos en estadística recomienda que las clases con frecuencias menores se agrupen en una clase agregada llamada “otro”. Las que presentan frecuencias de 5% o menos se tratan de esta manera.
2. La suma de las frecuencias en cualquier distribución de frecuencia es siempre igual al número de observaciones. La suma de las frecuencias relativas en cualquier distribución de frecuencia relativa es siempre igual a 1.00, y la de los porcentajes en una distribución de frecuencia porcentual es siempre igual a 100.

## Ejercicios

### Métodos

1. La respuesta a una pregunta tiene tres alternativas: A, B y C. Una muestra de 120 respuestas proporciona 60 A, 24 B y 36 C. Muestre las distribuciones de frecuencia y de frecuencia relativa.
2. Se tiene una distribución de frecuencia relativa parcial.

Clase	Frecuencia relativa
A	0.22
B	0.18
C	0.40
D	

- a) ¿Cuál es la frecuencia relativa de la clase D?
  - b) El tamaño de la muestra total es 200. ¿Cuál es la frecuencia de la clase D?
  - c) Muestre la distribución de frecuencia.
  - d) Muestre la distribución de frecuencia porcentual.
3. Un cuestionario proporciona 58 respuestas Sí, 42 No y 20 sin opinión.
    - a) En la elaboración de una gráfica circular, ¿cuántos grados mediría la sección del círculo que corresponde a las respuestas Sí?
    - b) ¿Cuántos grados mediría la sección del círculo que corresponde a las respuestas No?
    - c) Dibuje una gráfica circular.
    - d) Elabore una gráfica de barras.

## AUTO evaluación

### Aplicaciones

4. Los cuatro programas principales de televisión con mayor audiencia fueron *La ley y el orden (LyO)*, *CSI*, *Sin rastro* y *Esposas desesperadas* (ED) (Nielsen Media Research, 1 de enero de 2007). A continuación se proporcionan los datos que incluyen los programas preferidos en una muestra de 50 espectadores.

<i>ED</i>	<i>CSI</i>	<i>ED</i>	<i>CSI</i>	<i>LyO</i>
<i>Sin rastro</i>	<i>CSI</i>	<i>LyO</i>	<i>Sin rastro</i>	<i>CSI</i>
<i>CSI</i>	<i>ED</i>	<i>Sin rastro</i>	<i>CSI</i>	<i>ED</i>
<i>LyO</i>	<i>LyO</i>	<i>LyO</i>	<i>CSI</i>	<i>ED</i>
<i>CSI</i>	<i>ED</i>	<i>ED</i>	<i>LyO</i>	<i>CSI</i>
<i>ED</i>	<i>Sin rastro</i>	<i>CSI</i>	<i>Sin rastro</i>	<i>ED</i>
<i>ED</i>	<i>CSI</i>	<i>CSI</i>	<i>LyO</i>	<i>CSI</i>
<i>LyO</i>	<i>CSI</i>	<i>Sin rastro</i>	<i>Sin rastro</i>	<i>ED</i>
<i>LyO</i>	<i>CSI</i>	<i>CSI</i>	<i>CSI</i>	<i>ED</i>
<i>CSI</i>	<i>ED</i>	<i>Sin rastro</i>	<i>Sin rastro</i>	<i>LyO</i>

- a) ¿Estos datos son cualitativos o cuantitativos?
- b) Proporcione las distribuciones de frecuencia y frecuencia porcentual.
- c) Elabore una gráfica de barras y una gráfica circular.
- d) Según la muestra, ¿qué programa de televisión tiene la mayor audiencia? ¿Cuál es el segundo?

**WEB** archivo  
Names

5. En orden alfabético, los seis apellidos más comunes en Estados Unidos son Brown, Davis, Johnson, Jones, Smith y Williams (*The World Almanac*, 2006). Suponga que una muestra de 50 personas con uno de estos apellidos proporcionó los datos siguientes.

Brown	Williams	Williams	Williams	Brown
Smith	Jones	Smith	Johnson	Smith
Davis	Smith	Brown	Williams	Johnson
Johnson	Smith	Smith	Johnson	Brown
Williams	Davis	Johnson	Williams	Johnson
Williams	Johnson	Jones	Smith	Brown
Johnson	Smith	Smith	Brown	Jones
Jones	Jones	Smith	Smith	Davis
Davis	Jones	Williams	Davis	Smith
Jones	Johnson	Brown	Johnson	Davis

Resuma los datos mediante la elaboración de lo siguiente.

- a) Distribuciones de frecuencia relativa y frecuencia porcentual.
  - b) Una gráfica de barras.
  - c) Una gráfica circular.
  - d) Con base en estos datos, ¿cuáles son los tres apellidos más comunes?
6. El rating de audiencia televisiva de Nielsen Media Research mide el porcentaje de propietarios de un televisor que ven un programa en particular. El programa con mayor rating en la historia de la televisión estadounidense fue el último episodio especial de *M\*A\*S\*H*, transmitido el 28 de febrero de 1983. Un rating de 60.2 indicó que 60.2% de los televidentes lo vio. Nielsen Media Research proporcionó la lista de los 50 programas con mayor rating en la historia de la televisión (*The New York Times Almanac*, 2006). Los datos siguientes muestran la cadena que produjo cada uno de los 50 programas con mayor rating.

**WEB** archivo  
Networks

ABC	ABC	ABC	NBC	CBS
ABC	CBS	ABC	ABC	NBC
NBC	NBC	CBS	ABC	NBC
CBS	ABC	CBS	NBC	ABC
CBS	NBC	NBC	CBS	NBC
CBS	CBS	CBS	NBC	NBC
FOX	CBS	CBS	ABC	NBC
ABC	ABC	CBS	NBC	NBC
NBC	CBS	NBC	CBS	CBS
ABC	CBS	ABC	NBC	ABC

- a) Elabore una distribución de frecuencia, una distribución de frecuencia porcentual y una gráfica de barras para los datos.

**AUTO** evaluación

b) ¿Cuál o cuáles cadenas han transmitido los programas de televisión con los mayores ratings? Compare el desempeño de ABC, CBS y NBC.

7. Waterfront Steakhouse de Leverock, en Madeira Beach, Florida, utiliza un cuestionario para preguntar a sus clientes cómo califican el servicio, la calidad de los alimentos, los cocteles, los precios y la atmósfera del restaurante. Cada característica se califica en una escala de excepcional (O), muy bueno (V), bueno (G), promedio (A) y malo (P). Use la estadística descriptiva para resumir los datos siguientes recabados sobre la calidad de la comida. ¿Qué piensa sobre las calificaciones de la calidad en el restaurante?

G	O	V	G	A	O	V	O	V	G	O	V	A
V	O	P	V	O	G	A	O	O	O	G	O	V
V	A	G	O	V	P	V	O	O	G	O	O	V
O	G	A	O	V	O	O	G	V	A	G		

8. Los datos de una muestra de 55 miembros del Salón de la Fama de Béisbol en Cooperstown, Nueva York, se presentan enseguida. Cada observación indica la posición principal jugada por los famosos del Salón de la Fama: pitcher (P), catcher (H), primera base (1), segunda base (2), tercera base (3), parador en corto (S), jardinero izquierdo (L), jardinero central (C) y jardinero derecho (R).

L	P	C	H	2	P	R	1	S	S	1	L	P	R	P
P	P	P	R	C	S	L	R	P	C	C	P	P	R	P
2	3	P	H	L	P	1	C	P	P	P	S	1	L	R
R	1	2	H	S	3	H	2	L	P					

- a) Use las distribuciones de frecuencia y de frecuencia relativa para resumir los datos.  
 b) ¿Qué posición proporciona el mayor número de jugadores en el Salón de la Fama?  
 c) ¿Qué posición proporciona el menor número de jugadores?  
 d) ¿Qué posición de jardinero (L, C o R) proporciona la mayoría de jugadores en el Salón de la Fama?  
 e) Compare los jugadores de cuadro (1, 2, 3 y S) con los jugadores de campo (L, C y R).
9. El proyecto Tendencias demográficas y sociales del Pew Research Center encontró que 46% de los adultos estadounidenses preferiría vivir en un tipo distinto de comunidad que donde reside ahora (Pew Research Center, 29 de enero de 2009). La encuesta nacional preguntó a 2260 adultos: “¿En dónde vive ahora?” y “¿Cuál considera que es la comunidad ideal?” Las opciones de respuesta fueron ciudad (C), suburbio (S), ciudad pequeña (T) o comunidad rural (R). Una muestra de 100 personas se proporciona enseguida.

¿En dónde vive ahora?

S	T	R	C	R	R	T	C	S	T	C	S	C	S	T
S	S	C	S	S	T	T	C	C	S	T	C	S	T	C
T	R	S	S	T	C	S	C	T	C	T	C	T	C	R
C	C	R	T	C	S	S	T	S	C	C	C	R	S	C
S	S	C	C	S	C	R	T	T	T	C	R	T	C	R
C	T	R	R	C	T	C	C	R	T	T	R	S	R	T
T	S	S	S	S	S	C	C	R	T					

¿Cuál considera que es la comunidad ideal?

S	C	R	R	R	S	T	S	S	T	T	S	C	S	T
C	C	R	T	R	S	T	S	S	C	C	T	T	T	S
S	R	C	S	C	C	S	C	R	C	T	S	R	R	R
C	T	S	T	T	T	R	R	S	C	C	R	R	S	S
S	T	C	T	T	C	R	T	T	T	C	T	T	R	R
C	S	R	T	C	T	C	C	T	T	T	R	C	R	T
T	C	S	S	C	S	T	S	S	R					

- a) Proporcione una distribución de frecuencia porcentual para cada pregunta.  
 b) Trace una gráfica de barras para cada pregunta.  
 c) ¿En dónde vive ahora la mayoría de los adultos?  
 d) ¿Cuál considera la mayoría de los adultos que es la comunidad ideal?

**WEB** archivo  
LivingArea



**WEB** archivo  
FedBank

- e) ¿Qué cambios en las áreas habitables esperaría usted ver si las personas se mudan de donde viven actualmente a su comunidad ideal?
10. La *Financial Times/Harris* es una encuesta mensual en línea de adultos de seis países de Europa y Estados Unidos. La consulta realizada en enero de 2008 incluyó las respuestas de 1015 adultos. Una de las preguntas formuladas fue: “¿Cómo calificaría al Federal Bank en el manejo de los problemas crediticios de los mercados financieros?” Las respuestas posibles fueron excelente, bueno, justo, malo y terrible (sitio web Harris Interactive, enero de 2008). Las 1015 respuestas para esta pregunta pueden encontrarse en el archivo de datos llamado FedBank.
- Construya una distribución de frecuencia.
  - Prepare una distribución de frecuencia porcentual.
  - Elabore una gráfica de barras para la distribución de frecuencia porcentual.
  - Comente cómo piensan los adultos estadounidenses que el Federal Bank está manejando los problemas crediticios en los mercados financieros.
  - En España se preguntó a 1114 adultos “¿Cómo calificaría usted al Banco Central Europeo en el manejo de los problemas crediticios en los mercados financieros?” La distribución de frecuencia porcentual obtenida es la siguiente.

Calificación	Frecuencia porcentual
Excelente	0
Bueno	4
Justo	46
Malo	40
Terrible	10

Compare los resultados obtenidos en España con los resultados de Estados Unidos.

## 2.2

## Resumen de datos cuantitativos

### Distribución de frecuencia

**TABLA 2.4**

Duración de la auditoría de fin de año (en días)

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

Como se definió en la sección 2.1, una distribución de frecuencia es un resumen tabular de los datos que muestra el número (frecuencia) de elementos en cada una de varias clases que no se superponen. Esta definición es válida para los datos tanto cuantitativos como cualitativos. Sin embargo, con los datos cuantitativos debemos ser más cuidadosos al definir clases que no se superponen y que se utilizarán en la distribución de frecuencia.

Por ejemplo, considere los datos cuantitativos de la tabla 2.4. Éstos presentan el tiempo en días necesario para completar las auditorías de final de año para una muestra de 20 clientes de Sanderson and Clifford, una pequeña firma de contadores públicos. Los tres pasos necesarios para definir las clases de una distribución de frecuencia con los datos cuantitativos son los siguientes:

- Determine el número de clases que no se superponen.
- Defina el ancho de cada clase.
- Determine los límites de clase.

**WEB** archivo

**Audit**

Ahora se demostrarán estos pasos mediante el desarrollo de una distribución de frecuencia para los datos de duración de la auditoría de la tabla 2.4.

**Número de clases** Las clases se forman mediante la especificación de los rangos que se usarán para agrupar los datos. Como regla general, se recomienda utilizar entre 5 y 20 clases. En el caso de un número pequeño de elementos de datos se pueden utilizar cinco o seis clases para resumir los datos. Si se tienen muchos elementos, se requiere un número grande de clases. La idea es utilizar suficientes clases para mostrar la variación en los datos, pero no demasiadas si sólo se tienen algunos elementos. Dado que el número de elementos de datos en la tabla 2.4 es relativamente pequeño ( $n = 20$ ), se eligió elaborar una distribución de frecuencia con cinco clases.

*Asignar el mismo ancho a las clases reduce la posibilidad de interpretaciones inadecuadas de los usuarios.*

**Ancho de clase** El segundo paso en la elaboración de una distribución de frecuencia para datos cuantitativos es elegir el ancho de las clases. Como regla general, es recomendable que sea el mismo para todas. Por tanto, el número y el ancho de clase no son decisiones independientes. Un número grande de clases significa un ancho de clase menor, y viceversa. Para determinar un ancho de clase aproximado, primero se identifican los valores de datos mayores y menores. Luego, una vez especificado el número de clases deseado, se utiliza la expresión siguiente para determinar el ancho de clase aproximado.

$$\text{Ancho de clase aproximado} = \frac{\text{valor de datos mayor} - \text{valor de datos menor}}{\text{número de clases}} \quad (2.2)$$

El ancho de clase aproximado que se obtiene por la ecuación (2.2) se redondea a un valor más conveniente con base en la preferencia de la persona que elabora la distribución de frecuencia. Por ejemplo, un ancho de 9.28 podría redondearse a 10, sencillamente porque 10 es un ancho de clase más adecuado para la presentación de una distribución de frecuencia.

Para los datos que involucran la duración de la auditoría al final del año, el valor de datos mayor es 33 y el valor de datos menor es 12. Dado que se decidió resumir los datos en cinco clases, usando la ecuación (2.2) se obtiene un ancho de clase aproximado de  $(33 - 12)/5 = 4.2$ . Por tanto, se toma la decisión de redondear y usar un ancho de clase de cinco días en la distribución de frecuencia.

En la práctica, el número y ancho de clases apropiados se determinan por prueba y error. Una vez que se elige un número de clases determinado, la ecuación (2.2) se usa para encontrar el ancho de clase aproximado. El proceso se repite para los diferentes números de clases. En última instancia, el analista recurre a su juicio para determinar la combinación del número y ancho de clases que proporcionan la mejor distribución de frecuencia para resumir los datos.

En el caso de los datos de duración de la auditoría de la tabla 2.4, después de decidir utilizar cinco clases, cada una con un ancho de cinco días, el paso siguiente es especificar los límites de clase para cada una.

**Límites de clase** Deben elegirse de modo que cada elemento de datos pertenezca a una y sólo una de las clases. El *límite de clase inferior* identifica el valor de datos menor asignado a la clase. El *límite de clase superior* identifica el valor de datos mayor asignado a la clase. En la elaboración de distribuciones de frecuencia para datos cualitativos no se necesita especificar los límites de clase, debido a que cada elemento de datos corresponde de manera natural a una clase separada. Pero con los datos cuantitativos, como en el caso de la duración de las auditorías de la tabla 2.4, se necesitan los límites de clase para determinar a dónde pertenece cada valor de datos.

Utilizando los datos de duración de la auditoría de la tabla 2.4, se selecciona 10 días como el límite de clase inferior y 14 días como el límite de clase superior de la primera clase. Ésta se denota como 10–14 en la tabla 2.5. El valor de datos menor, 12, se incluye en la clase 10–14. Luego se selecciona 15 días como el límite de clase inferior y 19 días como límite superior de la clase siguiente. Enseguida se prosigue con la definición de los límites superior e inferior para obtener un total de cinco clases: 10–14, 15–19, 20–24, 25–29 y 30–34. El valor de datos mayor, 33, se incluye en la clase 30–34. La diferencia entre los límites inferiores de las clases adyacentes es el ancho de clase. Utilizando los primeros dos límites inferiores, 10 y 15, se observa que el ancho de clase es  $15 - 10 = 5$ .

Una vez determinados el número, ancho y límites de clase se obtiene una distribución de frecuencia mediante el conteo del número de valores de datos que pertenecen a cada clase. Por ejemplo, los datos de la tabla 2.4 muestran que cuatro valores, 12, 14, 14 y 13, pertenecen a la clase 10–14. Por tanto, la frecuencia de la clase 10–14 es 4. Al continuar con este proceso de conteo para las clases 15–19, 20–24, 25–29 y 30–34 se obtiene la distribución de frecuencia de la tabla 2.5. Esta distribución permite observar lo siguiente:

1. Las duraciones de las auditorías que ocurren con más frecuencia están en la clase 15–19 días. Ocho de las 20 duraciones de las auditorías pertenecen a esta clase.
2. Sólo una auditoría requirió 30 o más días.

Es posible formular otras conclusiones, dependiendo de los intereses de la persona que observa la distribución de frecuencia. Su utilidad estriba en que permite comprender los datos, lo que no se logra fácilmente con la simple observación de éstos en su forma desorganizada original.

*Ninguna frecuencia de datos es mejor para un conjunto de datos. Distintas personas pueden elaborar distribuciones de frecuencia diferentes, pero igualmente aceptables. La meta es mostrar el agrupamiento natural y la variación en los datos.*

**TABLA 2.5**

**Distribución de frecuencia para los datos de duración de la auditoría**

Duración de la auditoría (días)	Frecuencia
10–14	4
15–19	8
20–24	5
25–29	2
30–34	1
Total	20

**TABLA 2.6** Distribuciones de frecuencia relativa y frecuencia porcentual para los datos de duración de la auditoría

Duración de la auditoría (días)	Frecuencia relativa	Frecuencia porcentual
10–14	0.20	20
15–19	0.40	40
20–24	0.25	25
25–29	0.10	10
30–34	0.05	5
Total	1.00	100

**Punto medio de clase** En algunas aplicaciones se desea conocer los puntos medios de las clases en una distribución de frecuencia de datos cuantitativos. El **punto medio de clase** es el valor medio entre los límites de clase inferior y superior. En el caso de los datos de duración de la auditoría, los puntos medios de las cinco clases son 12, 17, 22, 27 y 32.

## Distribuciones de frecuencia relativa y frecuencia porcentual

Las distribuciones de frecuencia relativa y de frecuencia porcentual para los datos cuantitativos se definen de la misma manera que para los datos cualitativos. Primero, recuerde que la frecuencia relativa es la proporción de las observaciones que pertenecen a una clase. Si se tienen  $n$  observaciones:

$$\text{Frecuencia relativa de la clase} = \frac{\text{frecuencia de la clase}}{n}$$

La frecuencia porcentual de una clase es la frecuencia relativa multiplicada por 100.

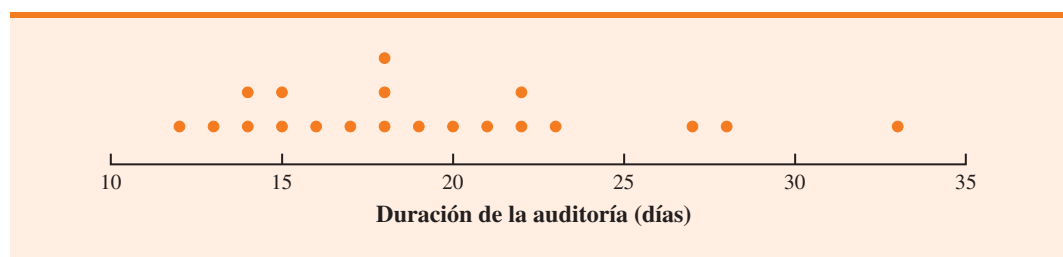
Con base en la frecuencia de las clases de la tabla 2.5 y el ancho  $n = 20$ , en la tabla 2.6 se muestra la distribución de frecuencia relativa y la distribución de frecuencia porcentual de los datos de duración de la auditoría. Observe que 0.40 de las auditorías, o 40%, requirió de 15 a 19 días, y sólo 0.05, o 5%, requirió 30 o más días. De nuevo, las interpretaciones y elementos de comprensión adicionales se obtienen usando la tabla 2.6.

## Diagrama de puntos

Uno de los resúmenes gráficos de datos más sencillos es el **diagrama de puntos**. El eje horizontal muestra el rango de los datos. Cada valor se representa por medio de un punto colocado sobre este eje. La figura 2.3 es el diagrama de puntos de los datos de duración de la auditoría de la tabla 2.4. Los tres puntos ubicados por encima de 18 sobre el eje horizontal indican que una duración de la auditoría de 18 días ocurrió tres veces. Los diagramas de puntos muestran los detalles de los datos y son útiles para comparar la distribución de los datos de dos o más variables.

## Histograma

El **histograma** es una presentación gráfica común de los datos cuantitativos. Este resumen gráfico se elabora para datos previamente resumidos, ya sea mediante una distribución de frecuencia,

**FIGURA 2.3** Diagrama de puntos para los datos de duración de la auditoría

de frecuencia relativa o de frecuencia porcentual. Para elaborar un histograma, la variable de interés se coloca sobre el eje horizontal y la frecuencia de que se trate, sobre el eje vertical. La frecuencia, la frecuencia relativa o la frecuencia porcentual de cada clase se indica con el trazo de un rectángulo, cuya base está determinada por los límites de clase sobre el eje horizontal, y cuya altura es la frecuencia, la frecuencia relativa o la frecuencia porcentual correspondiente.

La figura 2.4 es un histograma de los datos de duración de la auditoría. Observe que la clase con mayor frecuencia se muestra por medio del rectángulo que aparece encima de la clase 15–19 días. La altura del rectángulo indica que la frecuencia de esta clase es 8. Un histograma de la distribución de frecuencia relativa o porcentual de estos datos se ve igual que el de la figura 2.4, con la excepción de que el eje vertical se etiqueta con los valores de la frecuencia relativa o porcentual respectiva.

Como muestra esta figura, los rectángulos adyacentes de un histograma están en contacto uno con otro. A diferencia de una gráfica de barras, no hay una separación natural entre los rectángulos de las clases adyacentes. Este formato es la convención usual para los histogramas. Debido a que las clases de los datos de duración de la auditoría se establecen como 10–14, 15–19, 20–24, 25–29 y 30–34, parecería que se requieren espacios de una unidad entre las clases, de 14 a 15, de 19 a 20, de 24 a 25 y de 29 a 30. Sin embargo, estos espacios se eliminan cuando se elabora este tipo de gráfico. La eliminación de los espacios entre las clases de un histograma de los datos de duración de la auditoría sirve para mostrar que todos los valores entre el límite inferior de la primera clase y el límite superior de la última clase son posibles.

Uno de los usos más importantes del histograma es proporcionar información acerca de la forma de una distribución. La figura 2.5 presenta cuatro histogramas elaborados a partir de distribuciones de frecuencia relativa. El panel A lo muestra para un conjunto de datos moderadamente sesgado a la izquierda. Se dice que un histograma está sesgado a la izquierda si su cola se extiende más hacia esta dirección. Éste es típico para las calificaciones de exámenes: no hay calificaciones superiores a 100%, la mayoría es superior a 70%, y sólo algunas son realmente bajas. El panel B muestra el histograma para un conjunto de datos moderadamente sesgado a la derecha. Se dice que está sesgado a la derecha si su cola se extiende más hacia esta dirección. Un ejemplo de este tipo son los histogramas de datos, como los precios de las viviendas: algunas casas costosas crean el sesgo de la cola hacia la derecha.

El histograma C es simétrico. En este tipo, la cola izquierda imita la forma de la cola derecha. Los histogramas de datos que se encuentran en las aplicaciones nunca son perfectamente simétricos, pero en muchas aplicaciones lo son en forma aproximada. Con los datos de las calificaciones de la prueba SAT, las estaturas y los pesos de las personas, y así por el estilo, se generan histogramas aproximadamente simétricos. El histograma D está muy sesgado a la derecha. Éste se elaboró a partir de los datos sobre la cantidad de compras que los clientes realizaron a lo largo de un día en una tienda de ropa para dama. Con los datos de las aplicaciones a los negocios y la economía, a menudo se elaboran histogramas sesgados a la derecha. Por ejemplo,

**FIGURA 2.4** Histograma de los datos de duración de la auditoría

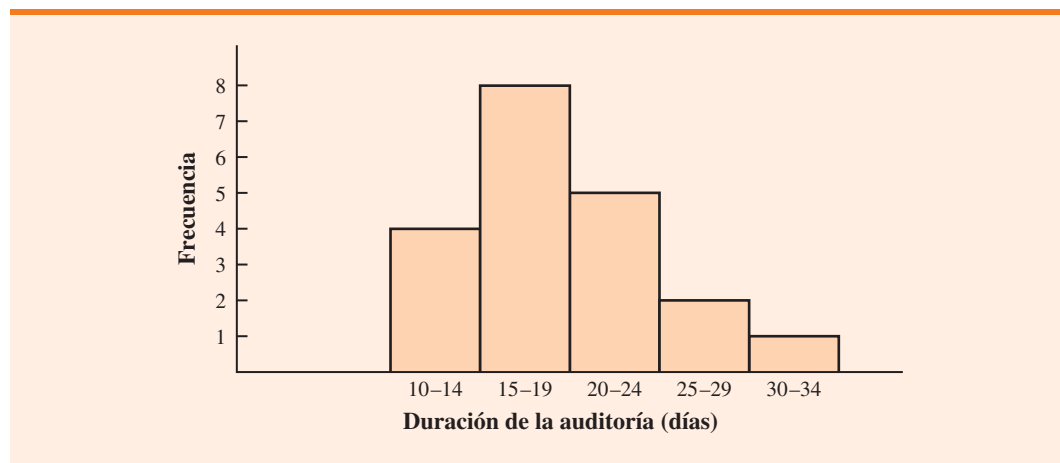
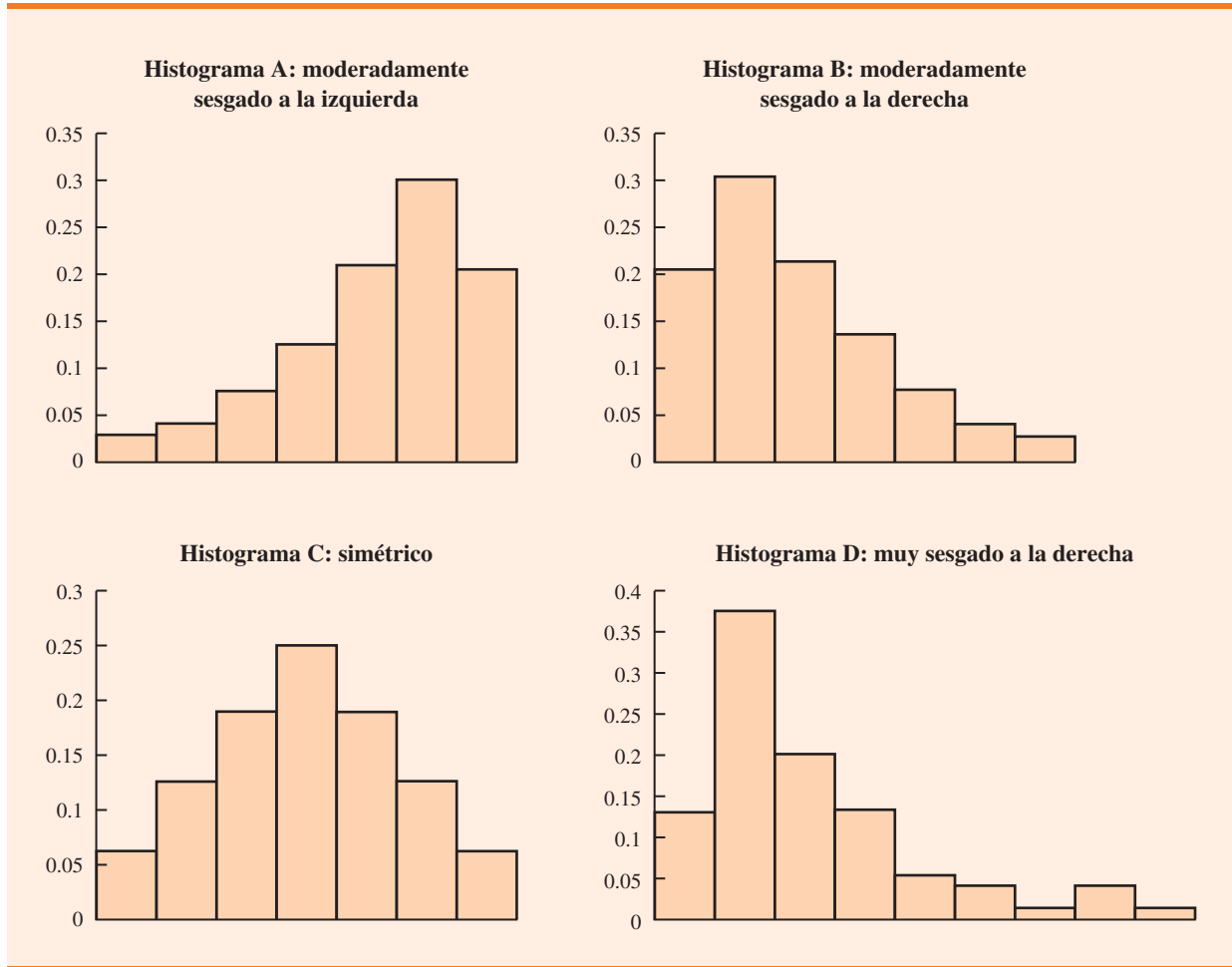


FIGURA 2.5 Histogramas que muestran diferentes niveles de sesgo



los datos sobre los precios de la vivienda, los sueldos, los montos de las compras, etc., con frecuencia generan gráficas de este tipo.

## Distribuciones acumuladas

Una variación de la distribución de frecuencia que proporciona otro resumen tabular de los datos cuantitativos es la **distribución de frecuencia acumulada**. Ésta utiliza el número, los anchos y los límites de clases desarrollados para la distribución de frecuencia. Sin embargo, en vez de indicar la frecuencia de cada clase, la distribución de frecuencia acumulada muestra el número de elementos de datos con valores *menores o iguales que el límite de clase superior de cada clase*. Las primeras dos columnas de la tabla 2.7 proporcionan la distribución de frecuencia acumulada de los datos de duración de la auditoría.

Para comprender cómo se determinan las frecuencias acumuladas, considere la clase con la descripción “Menos o igual que 24”. La frecuencia acumulada de esta clase es sencillamente la suma de las frecuencias de todas las clases con los valores de datos menores o iguales que 24. Para la distribución de frecuencia de la tabla 2.5, la suma de las frecuencias de las clases 10–14, 15–19 y 20–24 indica que  $4 + 8 + 5 = 17$  valores de datos son menores o iguales que 24. Por consiguiente, la frecuencia acumulada para esta clase es 17. Además, la distribución de frecuencia acumulada de la tabla 2.7 indica que se completaron cuatro auditorías en 14 días o menos y 19 auditorías en 29 días o menos.

**TABLA 2.7** Distribuciones de frecuencia acumulada, frecuencia relativa acumulada y frecuencia porcentual acumulada para los datos de duración de la auditoría

Duración de la auditoría (días)	Frecuencia acumulada	Frecuencia relativa acumulada	Frecuencia porcentual acumulada
Menos o igual que 14	4	0.20	20
Menos o igual que 19	12	0.60	60
Menos o igual que 24	17	0.85	85
Menos o igual que 29	19	0.95	95
Menos o igual que 34	20	1.00	100

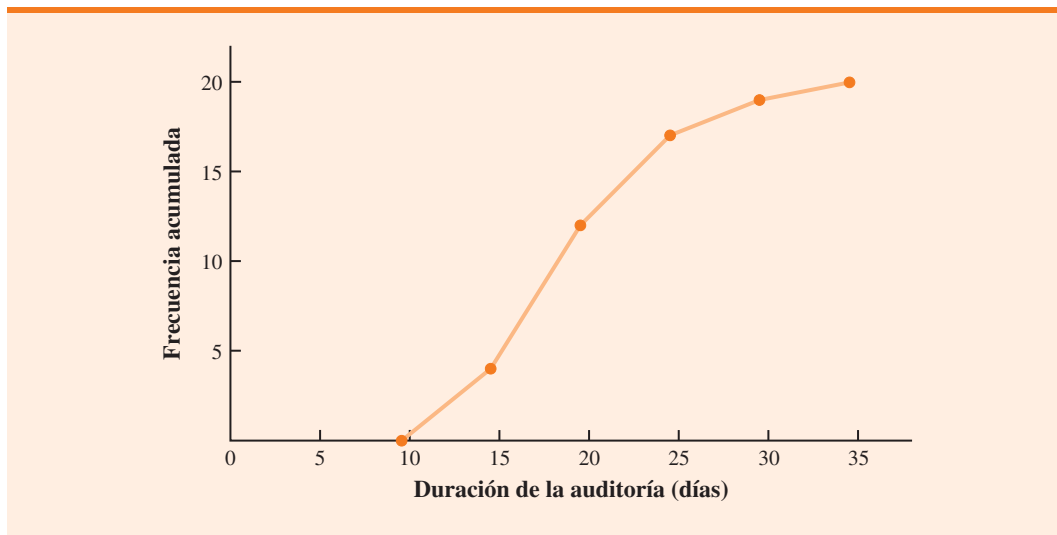
Para finalizar, observamos que una **distribución de frecuencia relativa acumulada** muestra la proporción de elementos de datos, y una **distribución de frecuencia porcentual acumulada** muestra el porcentaje de elementos de datos con valores inferiores o iguales al límite superior de cada clase. La distribución de frecuencia relativa acumulada se calcula con la suma de las frecuencias relativas en la distribución de frecuencia relativa o mediante la división de las frecuencias acumuladas entre el número total de elementos. Utilizando el último enfoque, las frecuencias relativas acumuladas de la columna 3 de la tabla 2.7 se obtuvieron al dividir las frecuencias acumuladas de la columna 2 entre el número total de elementos ( $n = 20$ ). Las frecuencias porcentuales acumuladas se calcularon de nuevo multiplicando las frecuencias relativas por 100. Las distribuciones de frecuencias relativa y porcentual acumuladas muestran que 0.85 de las auditorías, u 85%, se completó en 24 días o menos; 0.95 de las auditorías, o 95%, se completó en 29 días o menos, y así sucesivamente.

### Ojiva

La gráfica de una distribución acumulada, llamada **ojiva**, muestra los valores de datos sobre el eje horizontal, y ya sea las frecuencias acumuladas, las frecuencias relativas acumuladas o las frecuencias porcentuales acumuladas, sobre el eje vertical. La figura 2.6 ilustra la ojiva de las frecuencias acumuladas de los datos de duración de la auditoría de la tabla 2.7.

La ojiva se trazó marcando el punto correspondiente a la frecuencia acumulada de cada clase. Dado que las clases para los datos de duración de la auditoría son 10–14, 15–19, 20–24, etc., aparecen vacíos de una unidad entre 14 y 15, 19 y 20, etc. Éstos se eliminan al trazar pun-

**FIGURA 2.6** Ojiva de los datos de duración de la auditoría



tos a medio camino entre los límites de clase. Por tanto, 14.5 se usa para la clase 10–14; 19.5 para la clase 15–19, y así sucesivamente. La clase “Menos o igual que 14”, con una frecuencia acumulada de 4, se representa en la ojiva de la figura 2.6 por el punto localizado en 14.5 sobre el eje horizontal y 4 sobre el eje vertical. La clase “Menos o igual que 19”, con una frecuencia acumulada de 12, se representa por medio del punto localizado en 19.5 sobre el eje horizontal y 12 sobre el eje vertical. Note que se trazó otro punto en el extremo izquierdo de la ojiva. Este punto la inicia, lo cual indica que debajo de la clase 10–14 no hay ningún valor de datos. Este punto se localiza en 9.5 sobre el eje horizontal y 0 sobre el eje vertical. Los puntos trazados se conectan por medio de líneas rectas para completar la ojiva.

## NOTAS Y COMENTARIOS

- Una gráfica de barras y un histograma son en esencia lo mismo; ambos son representaciones gráficas de los datos de una distribución de frecuencia. Un histograma es sólo una gráfica de barras sin separación entre éstas. Para algunos datos cuantitativos discretos es apropiada una separación entre las barras. Considere, por ejemplo, el número de materias a las cuales se inscribe un estudiante universitario. Los datos sólo pueden asumir valores enteros. Los valores intermedios como 1.5, 2.73, etc., no son posibles. No obstante, con datos cuantitativos continuos, como la duración de la auditoría de la tabla 2.4, una separación entre las barras no es adecuada.
- Los valores apropiados para los límites de clase con datos cuantitativos dependen del nivel de precisión de éstos. Por ejemplo, con los datos de duración de la auditoría de la tabla 2.4 los límites empleados fueron valores enteros. Si se redondean a la décima más cercana de un día (p. ej., 12.3, 14.4, etc.), entonces los límites se establecerían en décimas de días. Por ejemplo, la primera clase sería 10.0–14.9. Si los datos se registraran a la centésima más cercana de un día (p. ej., 12.34, 14.45, etc.), los límites se fijarían en centésimas de días. Por ejemplo, la primera clase sería 10.00–14.99.
- Una clase de *extremo abierto* requiere sólo un límite de clase inferior o un límite de clase superior. Por ejemplo, en los datos de la tabla 2.4, suponga que dos de las auditorías han tardado 58 y 65 días. En vez de continuar con las clases de ancho 5 con las clases 35–39, 40–44, 45–49, etc., podríamos simplificar la frecuencia de distribución para mostrar una clase de extremo abierto de “35 o más”, la cual tendría una frecuencia de 2. Con más frecuencia la clase de extremo abierto aparece en el extremo superior de la distribución; a veces en el extremo inferior, y ocasionalmente en ambos extremos.
- La última entrada de una distribución de frecuencia acumulada es siempre igual al número total de observaciones, mientras que en una distribución de frecuencia relativa acumulada siempre es igual a 1.00, a la vez que la última entrada en una distribución de frecuencia porcentual acumulada siempre es igual a 100.

## Ejercicios

### Métodos

11. Considere los datos siguientes.

14	21	23	21	16
19	22	25	16	16
24	24	25	19	16
19	18	19	21	12
16	17	18	23	25
20	23	16	20	19
24	26	15	22	24
20	22	24	22	20

- Elabore una distribución de frecuencia usando las clases 12–14, 15–17, 18–20, 21–23 y 24–26.
- Desarrolle una distribución de frecuencia relativa y una distribución de frecuencia porcentual usando las clases del inciso a).

**WEB** archivo

Frequency

**AUTO** evaluación

12. Considere la distribución de frecuencia siguiente.

Clase	Frecuencia
10–19	10
20–29	14
30–39	17
40–49	7
50–59	2

Elabore una distribución de frecuencia acumulada y una distribución de frecuencia relativa acumulada.

13. Elabore un histograma y una ojiva para los datos del ejercicio 12.

14. Considere los datos siguientes.

8.9    10.2    11.5    7.8    10.0    12.2    13.5    14.1    10.0    12.2  
 6.8    9.5    11.5    11.2    14.9    7.5    10.0    6.0    15.8    11.5

- a) Elabore un diagrama de puntos.  
 b) Desarrolle una distribución de frecuencia.  
 c) Elabore una distribución de frecuencia porcentual.

**Aplicaciones****AUTO** evaluación

15. El personal del consultorio de un médico estudió los tiempos de espera de los pacientes que llegaron al consultorio con una solicitud de servicio de urgencia. Se recabaron los datos siguientes con los tiempos de espera en minutos durante un periodo de un mes.

2    5    10    12    4    4    5    17    11    8    9    8    12    21    6    8    7    13    18    3

Use las clases 0–4, 5–9, etc. para realizar lo siguiente:

- a) La distribución de frecuencia.  
 b) La distribución de frecuencia relativa.  
 c) La distribución de frecuencia acumulada.  
 d) La distribución de frecuencia relativa acumulada.  
 e) ¿Qué proporción de pacientes que necesita un servicio de emergencia espera 9 minutos o menos?  
 16. Una escasez de candidatos ha requerido que se paguen sueldos más altos y se ofrezcan beneficios adicionales para atraer y retener a los superintendentes de los distritos escolares. Los datos siguientes muestran el sueldo base anual (miles) para los superintendentes en 20 distritos del área más grande de Rochester, Nueva York (*The Rochester Democrat and Chronicle*, 10 de febrero de 2008).

187	184	174	185
175	172	202	197
165	208	215	164
162	172	182	156
172	175	170	183

Use las clases 150–159, 160–169, etc. para elaborar lo siguiente.

- a) La distribución de frecuencia.  
 b) La distribución de frecuencia porcentual.  
 c) La distribución de frecuencia porcentual acumulada.  
 d) Un histograma para el sueldo base anual.  
 e) ¿Los datos parecen estar sesgados? Explique por qué.  
 f) ¿Qué porcentaje de los superintendentes ganan más de \$200 000?  
 17. El promedio industrial Dow Jones (DJIA) sufrió una de sus poco frecuentes reorganizaciones de empresas cuando Cisco Systems y Travelers reemplazaron a General Motors y Citigroup (*The Wall Street Journal*, 8 de junio de 2009). En ese momento, los precios por acción para las 30 empresas en el DJIA eran los que se listan en el cuadro de la siguiente página.



**WEB** archivo  
DJIprices

Company	\$/Share	Company	\$/Share
3M	61	IBM	107
Alcoa	11	Intel	16
American Express	25	J.P. Morgan Chase	35
AT&T	24	Johnson & Johnson	56
Bank of America	12	Kraft Foods	27
Boeing	52	McDonald's	59
Caterpillar	38	Merck	26
Chevron	69	Microsoft	22
Cisco Systems	20	Pfizer	14
Coca-Cola	49	Procter & Gamble	53
DuPont	27	Travelers	43
ExxonMobil	72	United Technologies	56
General Electric	14	Verizon	29
Hewlett-Packard	37	Wal-Mart Stores	51
Home Depot	24	Walt Disney	25

- a) ¿Cuál es el precio por acción más alto? ¿Cuál es el precio por acción más bajo?
- b) Utilizando un ancho de clase de 10, elabore una distribución de frecuencia de los datos.
- c) Prepare un histograma e intérprete, incluyendo una discusión de su forma general, el rango del precio medio y el rango de precios más frecuente.
- d) Consulte *The Wall Street Journal* u otra publicación para encontrar el precio actual por acción de estas empresas. Elabore un histograma de los datos y comente cualquier cambio desde junio de 2009. ¿Qué empresa ha tenido el mayor incremento en el precio por acción? ¿Cuál ha tenido la disminución más grande?

**WEB** archivo  
Holiday

18. La investigación de NRF/BIG proporcionó los resultados de una encuesta de gastos vacacionales de los consumidores (*USA Today*, 20 de diciembre de 2005). Los datos siguientes indican la cantidad en dólares de gastos vacacionales para una muestra de 25 consumidores.

1 200	850	740	590	340
450	890	260	610	350
1 780	180	850	2 050	770
800	1 090	510	520	220
1 450	280	1 120	200	350

- a) ¿Cuál es el gasto vacacional más bajo? ¿Y el más alto?
  - b) Utilice un ancho de clase de \$250 para preparar una distribución de frecuencia y una distribución de frecuencia porcentual para los datos.
  - c) Elabore un histograma y comente la forma de la distribución.
  - d) ¿Qué observaciones puede plantear sobre los gastos vacacionales?
19. El correo electrónico no solicitado y la *spam* afectan la productividad de los empleados de oficina. Una encuesta de InsightExpress monitoreó a dichos empleados para determinar el tiempo improductivo por día dedicado a correo electrónico no solicitado y *spam* (*USA Today*, 13 de noviembre de 2003). Los datos siguientes presentan una muestra del tiempo en minutos dedicado a esta tarea.

2	4	8	4
8	1	2	32
12	1	5	7
5	5	3	4
24	19	4	14

Resuma los datos mediante la elaboración de lo siguiente:

- a) Una distribución de frecuencia (clases 1–5, 6–10, 11–15, 16–20, etcétera).
- b) Una distribución de frecuencia relativa.
- c) Una distribución de frecuencia acumulada.
- d) Una distribución de frecuencia relativa acumulada.
- e) Una ojiva.
- f) ¿Qué porcentaje de empleados de oficina pasó 5 minutos o menos en correo electrónico no solicitado o *spam*? ¿Qué porcentaje pasó más de 10 minutos al día en esta tarea?

20. *Golf Digest 50* publica una lista de los 50 golfistas profesionales con el ingreso anual total más alto. Éste representa la suma tanto de las ganancias en campo como de las obtenidas fuera de éste. Tiger Woods se clasificó en primer lugar con un ingreso total anual de \$122 millones. Sin embargo, casi \$100 millones de este total provinieron de actividades *fuera de campo*, como las promociones de productos y la apariencia personal. Los 10 golfistas profesionales con el ingreso fuera de campo más alto se muestran en la tabla siguiente (sitio web Golf Digest, febrero de 2008).

Nombre	Ingreso fuera de campo (miles de dólares)
Tiger Woods	99 800
Phil Mickelson	40 200
Arnold Palmer	29 500
Vijay Singh	25 250
Ernie Els	24 500
Greg Norman	24 000
Jack Nicklaus	20 750
Sergio Garcia	14 500
Michelle Wie	12 500
Jim Furyk	11 000

El ingreso fuera de campo de los 50 golfistas profesionales en el *Golf Digest 50* puede encontrarse en el sitio web del libro. Los datos se proporcionan en miles de dólares. Use las clases de 0–999, 5 000–9 999, 10 000–14 999, etc. para responder las preguntas siguientes. Incluya una clase de extremo abierto de 50 000 o más como la clase de ingresos más grande.

**WEB** archivo  
OffCourse

- Elabore una distribución de frecuencia y una distribución de frecuencia porcentual del ingreso anual fuera de campo de los 50 golfistas profesionales.
  - Elabore un histograma para estos datos.
  - Comente la forma de la distribución del ingreso fuera de campo.
  - ¿Cuál es la clase de ingresos fuera de campo más frecuente para los 50 golfistas profesionales? Usando sus resúmenes tabulares y gráficos, ¿qué observaciones adicionales puede hacer sobre el ingreso fuera de campo de estos deportistas?
21. El informe *Nielsen Home Technology Report* proporciona información sobre tecnología en el hogar y su uso. Los datos siguientes registran las horas de uso de computadoras personales durante una semana para una muestra de 50 personas.

**WEB** archivo  
Computer

4.1	1.5	10.4	5.9	3.4	5.7	1.6	6.1	3.0	3.7
3.1	4.8	2.0	14.8	5.4	4.2	3.9	4.1	11.1	3.5
4.1	4.1	8.8	5.6	4.3	3.3	7.1	10.3	6.2	7.6
10.8	2.8	9.5	12.9	12.1	0.7	4.0	9.2	4.4	5.7
7.2	6.1	5.7	5.9	4.7	3.9	3.7	3.1	6.1	3.1

Resuma los datos al elaborar lo siguiente:

- Una distribución de frecuencia (utilice un ancho de clase de 3 horas).
- Una distribución de frecuencia relativa.
- Un histograma.
- Una ojiva.
- Comente qué indican los datos sobre el uso de computadoras personales en casa.

## 2.3

### Análisis de datos exploratorios: el diagrama de tallo y hoja

Las técnicas del **análisis de datos exploratorio** consisten en una aritmética simple y gráficas fáciles de elaborar que pueden usarse para resumir los datos rápidamente. Una de ellas, conocida como **diagrama de tallo y hoja**, puede utilizarse para mostrar simultáneamente tanto la clasificación como la forma de un conjunto de datos.

TABLA 2.8 Número de preguntas respondidas correctamente en una prueba de aptitudes

112	72	69	97	107
73	92	76	86	73
126	128	118	127	124
82	104	132	134	83
92	108	96	100	92
115	76	91	102	81
95	141	81	80	106
84	119	113	98	75
68	98	115	106	95
100	85	94	106	119

**WEB** archivo  
ApTest

Para ilustrar el uso de un diagrama de tallo y hoja, considere los datos de la tabla 2.8. Éstos fueron obtenidos de una prueba de aptitudes de 150 preguntas a 50 personas entrevistadas recientemente para un puesto en Haskens Manufacturing. Los datos indican el número de preguntas respondidas correctamente.

Para desarrollar un diagrama de tallo y hoja, primero se colocan los dígitos principales de cada valor de datos a la izquierda de una línea vertical. A la derecha de dicha línea se registra el último dígito para cada valor de datos. Con base en la fila superior de los datos de la tabla 2.8 (112, 72, 69, 97 y 107), las primeras cinco entradas en la construcción de un diagrama de tallo y hoja serían como sigue:

```

6 | 9
7 | 2
8 |
9 | 7
10 | 7
11 | 2
12 |
13 |
14 |

```

Por ejemplo, el valor del dato 112 muestra los dígitos principales 11 a la izquierda de la línea y el último dígito 2 a la derecha. De modo parecido, el valor de datos 72 muestra el dígito principal 7 a la izquierda y el último dígito 2 a la derecha de la línea. Si se continúa colocando el último dígito de cada valor de datos en la línea que corresponde a sus dígitos principales, el resultado es el siguiente.

```

6 | 9 8
7 | 2 3 6 3 6 5
8 | 6 2 3 1 1 0 4 5
9 | 7 2 2 6 2 1 5 8 8 5 4
10 | 7 4 8 0 2 6 6 0 6
11 | 2 8 5 9 3 5 9
12 | 6 8 7 4
13 | 2 4
14 | 1

```

Con esta organización de los datos, la colocación de los dígitos en cada línea en el orden de clasificación es simple. Hacerlo así proporciona el diagrama de tallo y hoja mostrado aquí.

6	8	9
7	2	3 3 5 6 6
8	0	1 1 2 3 4 5 6
9	1	2 2 2 4 5 5 6 7 8 8
10	0	0 2 4 6 6 6 7 8
11	2	3 5 5 8 9 9
12	4	6 7 8
13	2	4
14	1	

Los números a la izquierda de la línea vertical (6, 7, 8, 9, 10, 11, 12, 13 y 14) forman el *tallo*, y cada dígito a la derecha de la línea es una *hoja*. Por ejemplo, considere la primera fila con un valor de tallo de 6 y las hojas de 8 y 9.

6 | 8 9

Esta fila indica que los dos valores de datos tienen un primer dígito 6. Las hojas muestran que los valores de datos son 68 y 69. De modo parecido, la segunda fila

7 | 2 3 3 5 6 6

indica que los seis valores de datos tienen un primer dígito 7. Las hojas muestran que los valores de datos son 72, 73, 73, 75, 76 y 76.

Para enfocarse en la forma indicada por el diagrama de tallo y hoja, usemos un rectángulo que contenga las hojas de cada tallo. Al hacerlo así se obtiene lo siguiente.

6	8 9
7	2 3 3 5 6 6
8	0 1 1 2 3 4 5 6
9	1 2 2 2 4 5 5 6 7 8 8
10	0 0 2 4 6 6 6 7 8
11	2 3 5 5 8 9 9
12	4 6 7 8
13	2 4
14	1

La rotación de esta página en sentido contrario a las manecillas del reloj sobre su lado proporciona una imagen de los datos similar a un histograma con las clases 60–69, 70–79, 80–89, etcétera.

Aunque puede parecer que el diagrama de tallo y hoja ofrece la misma información que un histograma, tiene dos ventajas principales.

1. El diagrama de tallo y hoja es más fácil de elaborar a mano.
2. Dentro de un intervalo de clase, proporciona más información que el histograma, debido a que el tallo y la hoja muestran los datos actuales.

Justamente como una distribución de frecuencia o un histograma no tienen un número absoluto de clases, ningún diagrama de tallo y hoja tiene un número absoluto de fila o tallos. Si creemos que nuestro diagrama original condensó demasiado los datos, podemos extenderlo fácilmente usando dos o más tallos para cada dígito principal. Por ejemplo, para usar dos tallos por cada

*En un diagrama de tallo y hoja extendido, siempre que un valor de tallo se establece dos veces, el primer valor corresponde a los valores de hoja de 0–4, y el segundo valor corresponde a los valores de hoja de 5–9.*

dígito principal, colocaríamos todos los valores de datos que terminan en 0, 1, 2, 3 y 4 en una fila, y todos los valores que terminan en 5, 6, 7, 8 y 9 en una segunda fila. El diagrama de tallo y hoja siguiente ilustra este enfoque.

6	8	9					
7	2	3	3				
7	5	6	6				
8	0	1	1	2	3	4	
8	5	6					
9	1	2	2	2	4		
9	5	5	6	7	8	8	
10	0	0	2	4			
10	6	6	6	7	8		
11	2	3					
11	5	5	8	9	9		
12	4						
12	6	7	8				
13	2	4					
13							
14	1						

Observe que los valores 72, 73 y 73 tienen hojas en el rango de 0–4 y se muestran con el primer valor de tallo de 7. Los valores 75, 76 y 76 tienen hojas en el rango de 5–9 y se registran con el segundo valor de tallo de 7. Este diagrama de tallo y hoja extendido es parecido a una distribución de frecuencia con intervalos de 65–69, 70–74, 75–79, etcétera.

El ejemplo anterior mostró un diagrama de tallo y hoja para los datos con hasta tres dígitos. Este tipo de diagramas para datos con más de tres dígitos es posible. Por ejemplo, considere los datos siguientes sobre el número de hamburguesas vendidas por un restaurante de comida rápida durante cada una de 15 semanas.

1565	1852	1644	1766	1888	1912	2044	1812
1790	1679	2008	1852	1967	1954	1733	

Un diagrama de tallo y hoja de estos datos se presenta a continuación.

Unidad de hoja = 10

15	6						
16	4	7					
17	3	6	9				
18	1	5	5	8			
19	1	5	6				
20	0	4					

*Un solo dígito se utiliza para definir cada hoja en un diagrama de tallo y hoja. La unidad de hoja indica cómo multiplicar los números de tallo y hoja con la finalidad de aproximar los datos originales. Las unidades de hoja pueden ser 100, 10, 1, 0.1, etcétera.*

Observe que se usa un solo dígito para definir cada hoja y que sólo los primeros tres dígitos de cada valor de datos se han utilizado para construir el diagrama. En la parte superior del diagrama hemos especificado que la unidad de hoja = 10. Para ilustrar cómo interpretar los valores, considere el primer tallo, 15, y su hoja asociada, 6. Al combinar estos números, obtenemos 156. Para reconstruir una aproximación de los valores de datos originales, debemos multiplicar este número por 10, el valor de la *unidad de hoja*. Por tanto,  $156 \times 10 = 1560$  es una aproximación del valor de datos original utilizado para construir el diagrama de tallo y hoja. Aunque no es posible reconstruir el valor de datos exacto a partir de este diagrama, la convención de manejar un sólo dígito para cada hoja permite que el diagrama se construya para datos que tienen muchos dígitos. Para diagramas donde la unidad de hoja no se muestra, se asume que esta unidad es igual a 1.

## Ejercicios

### Métodos

22. Elabore un diagrama de tallo y hoja para los datos siguientes.

70	72	75	64	58	83	80	82
76	75	68	65	57	78	85	72

### AUTO evaluación

23. Diseñe un diagrama de tallo y hoja para los datos siguientes.

11.3	9.6	10.4	7.5	8.3	10.5	10.0
9.3	8.1	7.7	7.5	8.4	6.3	8.8

24. Elabore un diagrama de tallo y hoja para los datos siguientes. Use una unidad de hoja de 10.

1 161	1 206	1 478	1 300	1 604	1 725	1 361	1 422
1 221	1 378	1 623	1 426	1 557	1 730	1 706	1 689

### Aplicaciones

### AUTO evaluación

25. Una psicóloga desarrolló una nueva prueba de inteligencia para adultos, la cual se aplicó a 20 individuos; se obtuvieron los datos siguientes.

114	99	131	124	117	102	106	127	119	115
98	104	144	151	132	106	125	122	118	118

Elabore un diagrama de tallo y hoja para los datos.

26. La Asociación Estadounidense de Inversionistas Individuales realiza una encuesta anual de corredores de descuento. Los precios siguientes corresponden a una muestra de 24 corredores o (brokers) (*AII Journal*, enero de 2003). Los dos tipos de negociaciones son una negociación asistida por un corredor de 100 acciones a \$50 por acción y una negociación en línea de 500 acciones a \$50 por acción.

Broker	Broker-Assisted 100 Shares at \$50/Share	Online 500 Shares at \$50/Share	Broker	Broker-Assisted 100 Shares at \$50/Share	Online 500 Shares at \$50/Share
Accutrade	30.00	29.95	Merrill Lynch Direct	50.00	29.95
Ameritrade	24.99	10.99	Muriel Siebert	45.00	14.95
Bank of America	54.00	24.95	NetVest	24.00	14.00
Brown & Co.	17.00	5.00	Recom Securities	35.00	12.95
Charles Schwab	55.00	29.95	Scottrade	17.00	7.00
CyberTrader	12.95	9.95	Sloan Securities	39.95	19.95
E*TRADE Securities	49.95	14.95	Strong Investments	55.00	24.95
First Discount	35.00	19.75	TD Waterhouse	45.00	17.95
Freedom Investments	25.00	15.00	T. Rowe Price	50.00	19.95
Harrisdirect	40.00	20.00	Vanguard	48.00	20.00
Investors National	39.00	62.50	Wall Street Discount	29.95	19.95
MB Trading	9.95	10.55	York Securities	40.00	36.00

### WEB archivo

#### Broker

- a) Redondee los precios comerciales al dólar más cercano y elabore un diagrama de tallo y hoja para 100 acciones a \$50 por acción. Comente qué aprendió acerca de los precios negociados asistidos por un corredor.
- b) Redondee los precios negociados al dólar más cercano y elabore un diagrama de tallo y hoja extendido para 500 acciones en línea a \$50 por acción. Comente qué aprendió sobre los precios negociados en línea.
27. La mayoría de las estaciones de esquí ofrece programas familiares que proporcionan instrucción de esquí y *snowboard* para niños. Las clases típicas constan de cuatro a seis horas en la nieve con un instructor certificado. A continuación se presenta la tarifa diaria para una lección de grupo de 15 estaciones (*The Wall Street Journal*, 20 de enero de 2006).

Estación	Ubicación	Tarifa diaria	Estación	Ubicación	Tarifa diaria
Beaver Creek	Colorado	\$137	Okemo	Vermont	\$ 86
Deer Valley	Utah	115	Park City	Utah	145
Diamond Peak	California	95	Butternut	Massachusetts	75
Heavenly	California	145	Steamboat	Colorado	98
Hunter	Nueva York	79	Stowe	Vermont	104
Mammoth	California	111	Sugar Bowl	California	100
Mount Sunapee	New Hampshire	96	Whistler-Blackcomb	British Columbia	104
Mount Bachelor	Oregon	83			

- a) Elabore un diagrama de tallo y hoja para los datos.  
 b) Interprete el diagrama en términos de qué indica la tarifa diaria para los programas de instrucción de esquí y *snowboard*.
28. En el minimaratón de Naples, Florida (13.1 millas), de 2004 se registraron 1 228 corredores (*Naples Daily News*, 17 de enero de 2004). La competencia se celebró en seis grupos de edades. Los datos siguientes muestran las edades de 40 individuos que participaron en la carrera.

**WEB** archivo  
 Marathon

49	33	40	37	56
44	46	57	55	32
50	52	43	64	40
46	24	30	37	43
31	43	50	36	61
27	44	35	31	43
52	43	66	31	50
72	26	59	21	47

- a) Muestre un diagrama de tallo y hoja.  
 b) ¿Qué grupo de edad tuvo el mayor número de corredores?  
 c) ¿Qué edad se registró con mayor frecuencia?  
 d) Un artículo de portada del *Naples Daily News* destacó el número de corredores de “veintitantos”. ¿Qué porcentaje de deportistas estaba en este grupo de edad? ¿Cuál se supone que fue el enfoque del artículo?

## 2.4

## Tabulaciones cruzadas y diagramas de dispersión

Las tabulaciones cruzadas y los diagramas de dispersión se usan para resumir datos de una manera en que revelan la relación entre dos variables.

Las tabulaciones cruzadas y los diagramas de dispersión se utilizan para resumir datos de una manera en que revelan la relación entre dos variables. Hasta ahora este capítulo se ha centrado en los métodos tabular y gráfico utilizados con el fin de resumir los datos para una *variable a la vez*. A menudo un gerente o quien toma decisiones requiere métodos de esa índole que le ayuden a comprender la *relación entre dos variables*. La tabulación cruzada y los diagramas de dispersión son dos métodos de este tipo.

### Tabulación cruzada

Una **tabulación cruzada** es un resumen de los datos para dos variables. Ilustremos su uso al considerar la aplicación siguiente con base en los datos de Zagat’s Restaurant Review. Se reunieron los datos sobre la calificación de calidad y precio de las comidas para una muestra de 300 restaurantes ubicados en el área de Los Ángeles. La tabla 2.9 muestra los datos para los primeros 10 negocios. Se informan los datos sobre la calificación de calidad de un restaurante y el precio de comidas típico. La calificación de calidad es una variable cualitativa con categorías de calificación de bueno, muy bueno y excelente. El precio de la comida es una variable cuantitativa que varía de \$10 a \$49.

Una tabulación cruzada de los datos para esta aplicación se muestra en la tabla 2.10. Las etiquetas del margen superior izquierdo definen las clases para las dos variables. En el margen izquierdo, las etiquetas de las filas buena, muy buena y excelente corresponden a las tres clases de la variable de calificación de la calidad. En el margen superior, las etiquetas de columna

**TABLA 2.9** Calificación de calidad y precio de la comida para 300 restaurantes de Los Ángeles

**WEB** archivo  
Restaurante

Restaurant	Quality Rating	Meal Price (\$)
1	Buena	18
2	Muy buena	22
3	Buena	28
4	Excelente	38
5	Muy buena	33
6	Buena	28
7	Muy buena	19
8	Muy buena	11
9	Muy buena	23
10	Buena	13
.	.	.
.	.	.
.	.	.

(\$10–19, \$20–29, \$30–39 y \$40–49) corresponden a las cuatro clases de la variable precio de la comida. Cada restaurante de la muestra exhibe una calificación de calidad y un precio de alimentos. Por tanto, cada uno se asocia con una celda que aparece en una de las filas y en una de las columnas de la tabulación cruzada. Por ejemplo, el restaurante 5 se identifica con una calificación de muy buena calidad y un precio de comida de \$33. Este negocio pertenece a la celda que intercepta la fila 2 y la columna 3 de la tabla 2.10. Al elaborar una tabulación cruzada, sencillamente contamos el número de restaurantes que pertenece a cada una de las celdas en la tabla.

Al revisar la tabla 2.10 observamos que el mayor número de restaurantes en la muestra (64) tiene calificación de muy buena calidad y un precio de comida en el rango de \$20–29. Sólo dos tienen una calificación de excelente y un precio de alimentos en el rango de \$10–19. Se pueden hacer interpretaciones parecidas de otras frecuencias. Además, observe que los márgenes derecho e inferior de la tabulación cruzada proporcionan por separado la distribución de frecuencia para la calificación de la calidad y el precio de la comida. De la distribución de frecuencia en el margen derecho, observamos que los datos sobre las calificaciones de calidad muestran 84 restaurantes buenos, 150 muy buenos y 66 excelentes. De modo parecido, el margen inferior ilustra la distribución de frecuencia para la variable del precio de la comida.

Al dividir los totales en el margen derecho de la tabulación cruzada entre el total para esa columna, se obtienen las distribuciones de frecuencia relativa y porcentual para la variable de calificación de la calidad.

Calificación de la calidad	Frecuencia relativa	Frecuencia porcentual
Buena	0.28	28
Muy buena	0.50	50
Excelente	0.22	22
<b>Total</b>	1.00	100

**TABLA 2.10** Tabulación cruzada de la calificación de calidad y el precio de la comida para 300 restaurantes de Los Ángeles

Calificación de calidad	Precio de la comida				Total
	\$10–19	\$20–29	\$30–39	\$40–49	
Buena	42	40	2	0	84
Muy buena	34	64	46	6	150
Excelente	2	14	28	22	66
<b>Total</b>	78	118	76	28	300



De la distribución de frecuencia porcentual, vemos que 28% de los restaurantes fue calificado como bueno, 50% muy bueno y 22% excelente.

Al dividir los totales en la fila inferior de la tabulación cruzada entre el total para esa fila se obtiene una distribución de frecuencia relativa y otra porcentual para la variable del precio de la comida.

Precio de la comida	Frecuencia relativa	Frecuencia porcentual
\$10–19	0.26	26
\$20–29	0.39	39
\$30–39	0.25	25
\$40–49	0.09	9
<b>Total</b>	1.00	100

Note que la suma de los valores de cada columna no se adiciona exactamente al total de la columna debido a que los valores se redondean. De la distribución de frecuencia porcentual vemos que 26% de los precios de la comida está en la clasificación más baja (\$10–19), 39% en la siguiente clase más alta, etcétera.

Las distribuciones de frecuencia relativa y porcentual elaboradas a partir de los márgenes proporcionan información acerca de cada una de las variables en forma individual, pero no arrojan luz sobre la relación entre las variables. El valor principal de una tabulación cruzada radica en los elementos que ofrece para comprender esta relación. Una revisión de la tabulación cruzada en la tabla 2.10 revela que los precios más altos de la comida se asocian con los restaurantes de mayor calidad, y los precios más bajos se ligan con los de menor calidad.

La conversión de las entradas en una tabulación cruzada en porcentajes de fila y de columna puede proporcionar más elementos para comprender la relación entre las dos variables. Para los porcentajes de fila, los resultados de dividir cada frecuencia entre su total de filas correspondiente en la tabla 2.10, se presentan en la tabla 2.11. Cada fila de esta tabla es una distribución de frecuencia porcentual del precio de la comida para una de las categorías de la calificación de calidad. De los restaurantes con la calificación de calidad menor (buena), los porcentajes mayores son para los menos caros (50% tiene precios de comida de \$10–19, y 47.6% de \$20–29). De los restaurantes con la calificación de calidad más alta (excelente), los porcentajes mayores son para los más caros (42.4% tiene precios de comida de \$30–39, y 33.4% de \$40–49). Por tanto, seguimos observando que los alimentos más caros se asocian con los restaurantes de mayor calidad.

La tabulación cruzada es de uso común al examinar la relación entre dos variables. En la práctica, los informes finales para muchos estudios estadísticos incluyen un número grande de tablas de este tipo. En la encuesta de restaurantes de Los Ángeles, la tabulación cruzada se basa en una variable cualitativa (calificación de la calidad) y una variable cuantitativa (precio de la comida). Las tabulaciones cruzadas también pueden elaborarse cuando ambas variables son cuantitativas y cuando ambas variables son cualitativas. Sin embargo, cuando se usan las cuantitativas, primero se deben crear clases para los valores de la variable. Por ejemplo, en el caso de los restaurantes agrupamos los precios de la comida en cuatro clases (\$10–19, \$20–29, \$30–39 y \$40–49).

**TABLA 2.11** Porcentajes de fila para cada categoría de calificación de la calidad

Calificación de calidad	Precio de la comida				Total
	\$10–19	\$20–29	\$30–39	\$40–49	
Buena	50.0	47.6	2.4	0.0	100
Muy buena	22.7	42.7	30.6	4.0	100
Excelente	3.0	21.2	42.4	33.4	100

## La paradoja de Simpson

Los datos en dos o más tabulaciones cruzadas se combinan o se agrupan con frecuencia para producir una tabulación cruzada de resumen que muestra cómo se relacionan las variables. En estos casos, debemos tener cuidado al formular una conclusión con base en los datos agregados, ya que ésta puede invertirse si estudiamos los datos no agregados. La revocación de las conclusiones basada en los datos agregados y no agregados se llama *paradoja de Simpson*. Para ilustrar la **paradoja de Simpson** considere un ejemplo que involucra el análisis de los veredictos de dos jueces en dos cortes distintas.

Los jueces Ron Luckett y Dennis Kendall presidieron los casos llevados en el tribunal de primera instancia y en el tribunal municipal en los tres años pasados. Algunos de los veredictos que emitieron se revocaron. En la mayoría de estos casos el tribunal de apelaciones ratificó las sentencias originales, pero en otros las revocó. Para cada juez se elaboró una tabulación cruzada con base en dos variables: el veredicto (ratificado o revocado) y el tipo de tribunal (primera instancia y municipal). Suponga que las dos tabulaciones cruzadas se combinaron al agregar los datos del tipo de tribunal. La tabulación cruzada agregada resultante contiene dos variables: el veredicto (ratificado o revocado) y el juez (Luckett o Kendall). Esta tabulación muestra el número de apelaciones en las cuales se ratificó la sentencia, el número y en las cuales se revocó para ambos jueces. La tabla siguiente ilustra estos resultados junto con los porcentajes de columna en el paréntesis al lado de cada valor.

Veredicto	Juez		Total
	Luckett	Kendall	
Ratificado	129 (86%)	110 (88%)	239
Revocado	21 (14%)	15 (12%)	36
<b>Total (%)</b>	150 (100%)	125 (100%)	275

Una revisión de los porcentajes de la columna muestra que 86% de los veredictos se ratificó para el juez Luckett, mientras que 88% se ratificó para el juez Kendall. A partir de esta tabulación cruzada agregada concluimos que Kendall está efectuando un mejor trabajo debido a que un mayor porcentaje de sus sentencias ha sido ratificado.

Las tabulaciones cruzadas no agregadas muestran los casos juzgados por Luckett y Kendall en cada tribunal; los porcentajes de columna se registran en el paréntesis al lado de cada valor.

Veredicto	Juez Luckett			Veredicto	Juez Kendall		
	Tribunal de primera instancia	Tribunal municipal	Total		Tribunal de primera instancia	Tribunal municipal	Total
Ratificado	29 (91%)	100 (85%)	129	Ratificado	90 (90%)	20 (80%)	110
Revocado	3 (9%)	18 (15%)	21	Revocado	10 (10%)	5 (20%)	15
<b>Total (%)</b>	32 (100%)	118 (100%)	150	<b>Total (%)</b>	100 (100%)	25 (100%)	125

A partir de la tabulación cruzada y los porcentajes de columna para el juez Luckett, observamos que los veredictos se ratificaron en 91% de los casos del tribunal de primera instancia y en 85% de los casos del tribunal municipal. De la tabulación cruzada y los porcentajes de columna para Kendall, los veredictos se mantienen en 90% de los casos del tribunal de primera instancia y en 80% de los correspondientes al tribunal municipal. Por tanto, cuando desagregamos los datos, observamos que Luckett tiene un mejor registro debido a que el mayor porcentaje de sus veredictos se mantiene en ambos tribunales. Este resultado contradice la conclusión a la que llegamos con la tabulación cruzada de los datos agregados que mostraron que Kendall tenía un mejor registro. Esta revocación de las conclusiones con base en los datos agregados y desagregados ilustra la paradoja de Simpson.

La tabulación cruzada original se obtuvo al agregar los datos en las tabulaciones cruzadas separadas para los dos tribunales. Note que para ambos jueces el porcentaje de apelaciones que dio como resultado revocaciones fue mucho mayor en el tribunal municipal que en el tribunal de primera instancia. Debido a que Luckett procesó un porcentaje mucho más alto de sus casos en el tribunal municipal, los datos agregados favorecieron al juez Kendall. No obstante, cuando miramos las tabulaciones cruzadas para los dos tribunales por separado, Luckett muestra el mejor registro. Por consiguiente, para la tabulación cruzada original, el *tipo de tribunal* es una variable oculta que no puede ignorarse cuando se evalúan los registros de los dos jueces.

Dada la posibilidad de la paradoja de Simpson, dése cuenta de que la conclusión o interpretación puede revocarse dependiendo de si usted está viendo datos de tabulación cruzada desagregados o agregados. Antes de formular una conclusión, tal vez quiera investigar si la forma agregada o desagregada de la tabulación proporciona la mejor comprensión y conclusión. En particular, cuando la tabulación cruzada involucra datos agregados, usted debe investigar si una variable oculta podría afectar los resultados, ya que las tabulaciones separadas o desagregadas proporcionan una comprensión y una conclusión diferentes y posiblemente mejores.

### Diagrama de dispersión y línea de tendencia

Un **diagrama de dispersión** es una presentación gráfica de la relación entre dos variables cuantitativas, y una **línea de tendencia** es aquella que proporciona una aproximación de la relación. Como ejemplo, considere la relación publicidad/ventas para una tienda de estéreo y equipos de sonido en San Francisco. Durante los tres meses pasados, en 10 ocasiones la tienda usó los comerciales de televisión de fin de semana para promover las ventas en sus establecimientos. Los gerentes quieren investigar si existe una relación entre el número de comerciales transmitidos y las ventas en la tienda durante la semana siguiente. Los datos muestrales de las 10 semanas con las ventas en cientos de dólares se registran en la tabla 2.12.

La figura 2.7 ilustra el diagrama de dispersión y la línea de tendencia<sup>1</sup> para los datos de la tabla 2.12. El número de comerciales ( $x$ ) se observa en el eje horizontal y las ventas ( $y$ ) en el eje vertical. Para la semana 1,  $x = 2$  y  $y = 50$ . Un punto con esas coordenadas se traza en el diagrama de dispersión. Puntos similares se trazan en las otras nueve semanas. Observe que en dos semanas se transmitió un comercial, durante dos de las semanas se mostraron dos comerciales, etcétera.

El diagrama de dispersión completo de la figura 2.7 indica una relación positiva entre el número de comerciales y las ventas. Las ventas más altas se asocian con un número mayor de anuncios publicitarios. La relación no es perfecta, ya que no todos los puntos están en una línea recta; sin embargo, el patrón general de los puntos y la línea de tendencia sugieren que en general la relación es positiva.

**TABLA 2.12** Datos muestrales para la tienda de estéreo y equipos de sonido

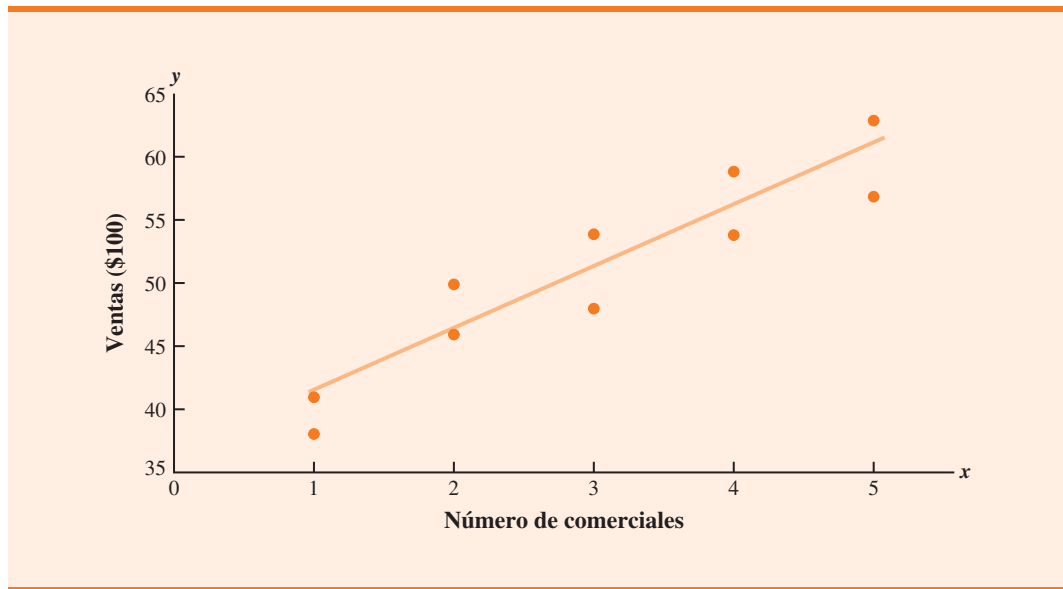
Week	Number of Commercials	Sales (\$100s)
	$x$	$y$
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

**WEB** archivo

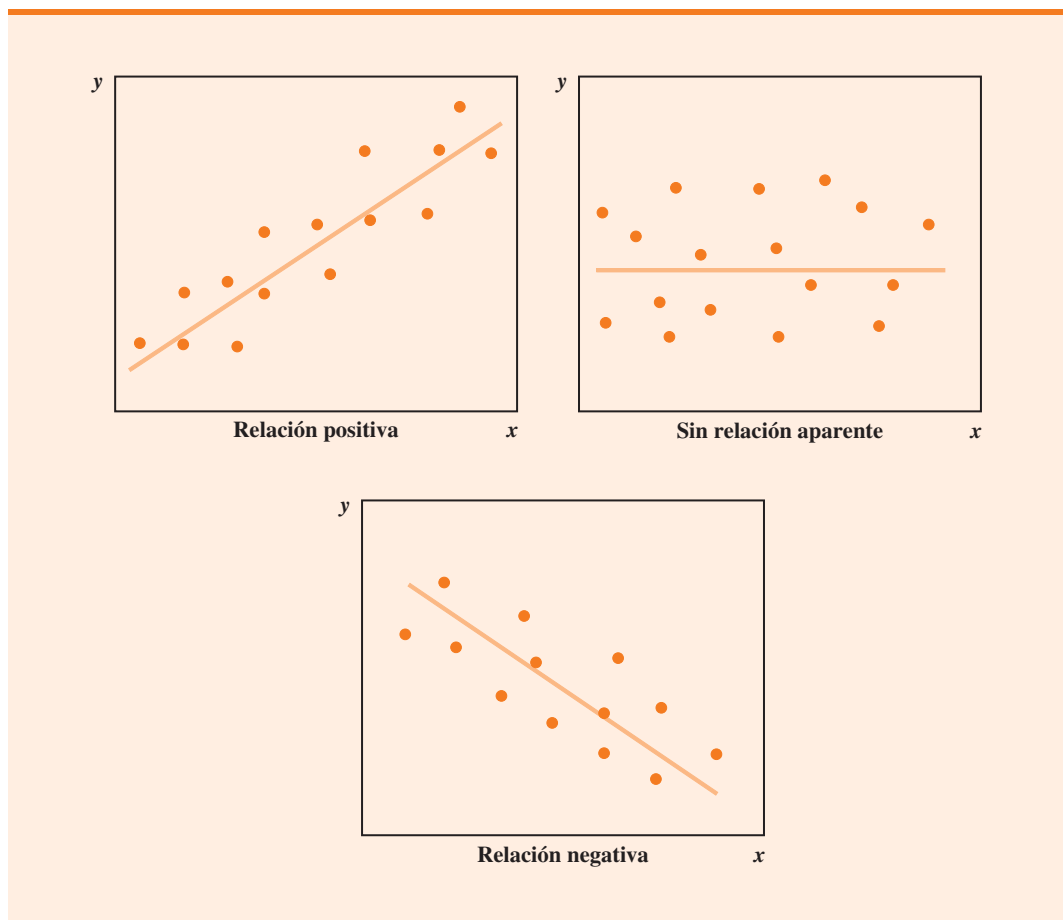
Stereo

<sup>1</sup> La ecuación de la línea de tendencia es  $y = 36.15 + 4.95x$ . La pendiente de la línea de tendencia es 4.95 y el intercepto en  $y$  (el punto donde la línea intercepta el eje  $y$ ) es 36.15. Comentaremos con detalle la interpretación de la pendiente y el intercepto en  $y$  para una línea de tendencia lineal en el capítulo 14, cuando estudiemos la regresión lineal simple.

**FIGURA 2.7** Diagrama de dispersión y línea de tendencia para la tienda de estéreos y equipos de sonido



**FIGURA 2.8** Tipos de relaciones representados por los diagramas de dispersión



Algunos patrones generales de diagramas de dispersión y los tipos de relación que sugieren se muestran en la figura 2.8. El panel superior izquierdo representa una relación positiva parecida a la del ejemplo del número de comerciales y ventas. En el panel superior derecho, el diagrama de dispersión no muestra una relación aparente entre las variables. El panel inferior representa una relación negativa donde  $y$  tiende a disminuir a medida que  $x$  aumenta.

## Ejercicios

### Métodos

29. Los datos siguientes corresponden a 30 observaciones que involucran dos variables cualitativas,  $x$  y  $y$ . Las categorías para  $x$  son A, B y C; las categorías para  $y$  son 1 y 2.

**AUTO** evaluación

**WEB** archivo

Crosstab

Observation	$x$	$y$	Observation	$x$	$y$
1	A	1	16	B	2
2	B	1	17	C	1
3	B	1	18	B	1
4	C	2	19	C	1
5	B	1	20	B	1
6	C	2	21	C	2
7	B	1	22	B	1
8	C	2	23	C	2
9	A	1	24	A	1
10	B	1	25	B	1
11	A	1	26	C	2
12	B	1	27	C	2
13	C	2	28	A	1
14	C	2	29	B	1
15	C	2	30	B	2

- Elabore una tabulación cruzada para los datos, con  $x$  como la variable de la fila y  $y$  como la variable de la columna.
- Calcule los porcentajes de la fila.
- Calcule los porcentajes de la columna.
- ¿Cuál es la relación, si existe, entre  $x$  y  $y$ .

**AUTO** evaluación

**WEB** archivo

Scatter

Observation	$x$	$y$	Observation	$x$	$y$
1	-22	22	11	-37	48
2	-33	49	12	34	-29
3	2	8	13	9	-18
4	29	-16	14	-33	31
5	-13	10	15	20	-16
6	21	-28	16	-3	14
7	-13	27	17	-15	18
8	-23	35	18	12	17
9	14	-5	19	-20	-11
10	3	-3	20	-7	-22

- Elabore un diagrama de dispersión para la relación entre  $x$  y  $y$ .
- ¿Cuál es la relación, si existe, entre  $x$  y  $y$ ?

## Aplicaciones

31. La tabulación cruzada siguiente muestra el ingreso familiar por nivel educativo del jefe de familia (*Statistical Abstract of the United States, 2008*).

Nivel educativo	Ingresos familiares (\$1 000)					Total
	Menos de 25	25.0–49.9	50.0–74.9	75.0–99.9	100 o más	
Sin educación media	4 207	3 459	1 389	539	367	9 961
Educación media	4 917	6 850	5 027	2 637	2 668	22 099
Educación superior inconclusa	2 807	5 258	4 678	3 250	4 074	20 067
Educación superior	885	2 094	2 848	2 581	5 379	13 787
Maestría o doctorado	290	829	1 274	1 241	4 188	7 822
<b>Total</b>	13 106	18 490	15 216	10 248	16 676	73 736

- a) Calcule los porcentajes de la fila e identifique la distribución de frecuencia porcentual para las familias donde el jefe tiene educación media y donde posee educación superior.
- b) ¿Qué porcentaje de familias dirigidas por una persona con educación media gana \$75 000 o más? ¿Qué porcentaje de familias encabezadas por una persona con educación superior gana \$75 000 o más?
- c) Elabore histogramas de frecuencia porcentual de los ingresos de familias dirigidas por personas con educación media y con educación superior. ¿La relación entre los ingresos familiares y el nivel educativo es evidente?
32. Consulte de nuevo la tabulación cruzada de los ingresos familiares por nivel educativo que se muestra en el ejercicio 31.
- a) Calcule los porcentajes de columna e identifique la distribución de frecuencia porcentual mostrada. ¿Qué proporción de los jefes de familia no cuenta con educación media?
- b) ¿Qué porcentaje de las familias que ganan \$100 000 o más están encabezadas por una persona que cuenta con una maestría o un doctorado? ¿Qué porcentaje de las familias dirigidas por una persona con alguno de esos grados gana más de \$100 000? ¿Por qué difieren estos dos porcentajes?
- c) Compare la distribución de frecuencia porcentual de las familias que ganan “menos de 25”, “100 o más” y del “total”. Comente la relación entre el ingreso familiar y el nivel educativo del jefe de familia.
33. Recientemente, la gerencia de Oak Tree Golf Course ha recibido algunas quejas sobre el estado de los *greens*. Varios jugadores se quejaron de que son muy rápidos. En vez de reaccionar a los comentarios de unos cuantos, la asociación de golf aplicó una encuesta a 100 hombres y 100 mujeres golfistas. Los resultados se resumen a continuación.

Golfistas hombres			Golfistas mujeres		
Handicap	Estado de los greens		Handicap	Estado de los greens	
	Demasiado rápidos	Bien		Demasiado rápidos	Bien
Menos de 15	10	40	Menos de 15	1	9
15 o más	25	25	15 o más	39	51

- a) Combine en una estas dos tablas de contingencia con las etiquetas de fila *hombres* y *mujeres*, y las etiquetas de columna *demasiado rápidos* y *bien*. ¿Qué grupo muestra el porcentaje más alto que afirma que los *greens* son demasiado rápidos?

- b) Consulte las tabulaciones cruzadas iniciales. Para los jugadores de handicap bajo (los mejores), ¿cuál grupo (hombres o mujeres) muestra el porcentaje más alto que afirma que los *greens* son demasiado rápidos?
- c) Consulte las tabulaciones cruzadas iniciales. Para los jugadores de handicap alto, ¿cuál grupo (hombres o mujeres) muestra el porcentaje más alto que afirma que los *greens* son demasiado rápidos?
- d) ¿Qué conclusiones puede formular acerca de las preferencias de los hombres y las mujeres respecto de la rapidez de los *greens*? ¿Las conclusiones del inciso a) en comparación con las de los incisos b) y c) son consistentes? Explique cualquier incongruencia aparente.
34. La tabla 2.13 de la siguiente página muestra un conjunto de datos con información para 45 fondos de inversión que son parte del *Morningstar Funds500* de 2008. El conjunto de datos incluye las cinco variables siguientes:
- Tipo de fondo.* El tipo de fondo, etiquetado como DE (capital nacional), IE (capital internacional) y FI (renta fija)
- Valor neto de los activos.* El precio de cierre por acción
- Rendimiento promedio a 5 años (%).* El rendimiento promedio anual para el fondo durante los cinco años anteriores
- Razón de gastos (%).* El porcentaje de activos deducido cada año fiscal para los gastos de fondos
- Calificación Morningstar.* La calificación, con estrellas, del riesgo ajustada para cada fondo; las calificaciones de Morningstar varían de una baja de 1 estrella (1-star) a una alta de 5 estrellas (5-stars)
- a) Elabore una tabulación cruzada de los datos sobre el tipo de fondo (filas) y el rendimiento promedio anual durante los cinco años anteriores (columnas). Utilice clases de 0–9.99; 10–19.99; 20–29.99; 30–39.99; 40–49.99, y 50–59.99 para el rendimiento promedio de 5 años (%).
- b) Prepare una distribución de frecuencia para los datos sobre el tipo de fondo.
- c) Elabore una distribución de frecuencia para los datos sobre el rendimiento promedio de 5 años (%).
- d) ¿Cómo ayudó la tabulación cruzada para la preparación de la distribución de frecuencias en los incisos b) y c)?
- e) ¿Qué conclusiones puede formular sobre el tipo de fondo y el rendimiento promedio sobre los 5 años anteriores?
35. Consulte los datos de la tabla 2.13.
- a) Elabore una tabulación cruzada de los datos sobre el tipo de fondo (filas) y la razón de gastos (columnas). Use las clases de 0.25–0.49; 0.50–0.74; 0.75–0.99; 1.00–1.24, y 1.25–1.49 para la razón de gastos (%).
- b) Prepare una distribución de frecuencia porcentual para la razón de gastos (%).
- c) ¿Qué conclusiones puede formular acerca del tipo de fondo y la razón de gastos?
36. Consulte los datos de la tabla 2.13.
- a) Elabore un diagrama de dispersión con un rendimiento promedio de cinco años (%) sobre el eje horizontal y el valor neto de los activos (\$) sobre el eje vertical.
- b) Comente la relación, si existe, entre las variables.
37. La *Guía de economía de combustible* del Departamento de Energía de Estados Unidos proporciona datos sobre la eficiencia de combustible para automóviles y camiones (sitio web Fuel Economy, 22 de febrero de 2008). Una porción de los datos para 311 automóviles compactos, medianos y grandes se muestra en la tabla 2.14. El conjunto de datos contiene las variables siguientes:
- Tamaño: compacto, mediano y grande
- Desplazamiento: tamaño del motor en litros
- Cilindros: número de cilindros en el motor
- Tracción: delantera (F), trasera (R) y en las cuatro llantas (4)
- Tipo de combustible: premium (P) o regular (R)
- Mi/gal en ciudad: calificación de la eficiencia del combustible para uso del automóvil en la ciudad en términos de millas por galón
- Mi/gal en autopista: calificación de la eficiencia del combustible para uso del automóvil en autopista en términos de millas por galón

El conjunto de datos completo está contenido en el archivo llamado FuelData08.

- Prepare una tabulación cruzada de los datos de tamaño (filas) y las millas por galón en autopista (columnas). Use las clases 15–19, 20–24, 25–29, 30–34 y 35–39 para mi/gal en autopista.
- Comente la relación entre el tamaño y las millas por galón en autopista.

**TABLA 2.13** Datos financieros para una muestra de 45 fondos de inversión

Fund Name	Fund Type	Net Asset Value (\$)	5-Year Average Return (%)	Expense Ratio (%)	Morningstar Rank
Amer Cent Inc & Growth Inv	DE	28.88	12.39	0.67	2-Star
American Century Intl. Disc	IE	14.37	30.53	1.41	3-Star
American Century Tax-Free Bond	FI	10.73	3.34	0.49	4-Star
American Century Ultra	DE	24.94	10.88	0.99	3-Star
Ariel	DE	46.39	11.32	1.03	2-Star
Artisan Intl Val	IE	25.52	24.95	1.23	3-Star
Artisan Small Cap	DE	16.92	15.67	1.18	3-Star
Baron Asset	DE	50.67	16.77	1.31	5-Star
Brandywine	DE	36.58	18.14	1.08	4-Star
Brown Cap Small	DE	35.73	15.85	1.20	4-Star
Buffalo Mid Cap	DE	15.29	17.25	1.02	3-Star
Delafield	DE	24.32	17.77	1.32	4-Star
DFA U.S. Micro Cap	DE	13.47	17.23	0.53	3-Star
Dodge & Cox Income	FI	12.51	4.31	0.44	4-Star
Fairholme	DE	31.86	18.23	1.00	5-Star
Fidelity Contrafund	DE	73.11	17.99	0.89	5-Star
Fidelity Municipal Income	FI	12.58	4.41	0.45	5-Star
Fidelity Overseas	IE	48.39	23.46	0.90	4-Star
Fidelity Sel Electronics	DE	45.60	13.50	0.89	3-Star
Fidelity Sh-Term Bond	FI	8.60	2.76	0.45	3-Star
Fidelity	DE	39.85	14.40	0.56	4-Star
FPA New Income	FI	10.95	4.63	0.62	3-Star
Gabelli Asset AAA	DE	49.81	16.70	1.36	4-Star
Greenspring	DE	23.59	12.46	1.07	3-Star
Janus	DE	32.26	12.81	0.90	3-Star
Janus Worldwide	IE	54.83	12.31	0.86	2-Star
Kalmar Gr Val Sm Cp	DE	15.30	15.31	1.32	3-Star
Managers Freemont Bond	FI	10.56	5.14	0.60	5-Star
Marsico 21st Century	DE	17.44	15.16	1.31	5-Star
Mathews Pacific Tiger	IE	27.86	32.70	1.16	3-Star
Meridan Value	DE	31.92	15.33	1.08	4-Star
Oakmark I	DE	40.37	9.51	1.05	2-Star
PIMCO Emerg Mkts Bd D	FI	10.68	13.57	1.25	3-Star
RS Value A	DE	26.27	23.68	1.36	4-Star
T. Rowe Price Latin Am.	IE	53.89	51.10	1.24	4-Star
T. Rowe Price Mid Val	DE	22.46	16.91	0.80	4-Star
Templeton Growth A	IE	24.07	15.91	1.01	3-Star
Thornburg Value A	DE	37.53	15.46	1.27	4-Star
USAA Income	FI	12.10	4.31	0.62	3-Star
Vanguard Equity-Inc	DE	24.42	13.41	0.29	4-Star
Vanguard Global Equity	IE	23.71	21.77	0.64	5-Star
Vanguard GNMA	FI	10.37	4.25	0.21	5-Star
Vanguard Sht-Tm TE	FI	15.68	2.37	0.16	3-Star
Vanguard Sm Cp Idx	DE	32.58	17.01	0.23	3-Star
Wasatch Sm Cp Growth	DE	35.41	13.98	1.19	4-Star



TABLA 2.14 Datos de la eficiencia de combustible para 311 automóviles

Car	Size	Displacement	Cylinders	Drive	Fuel Type	City MPG	Hwy MPG
1	Compacto	3.1	6	4	P	15	25
2	Compacto	3.1	6	4	P	17	25
3	Compacto	3.0	6	4	P	17	25
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
161	Mediano	2.4	4	F	R	22	30
162	Mediano	2.0	4	F	P	19	29
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
310	Grande	3.0	6	F	R	17	25
311	Grande	3.0	6	F	R	18	25

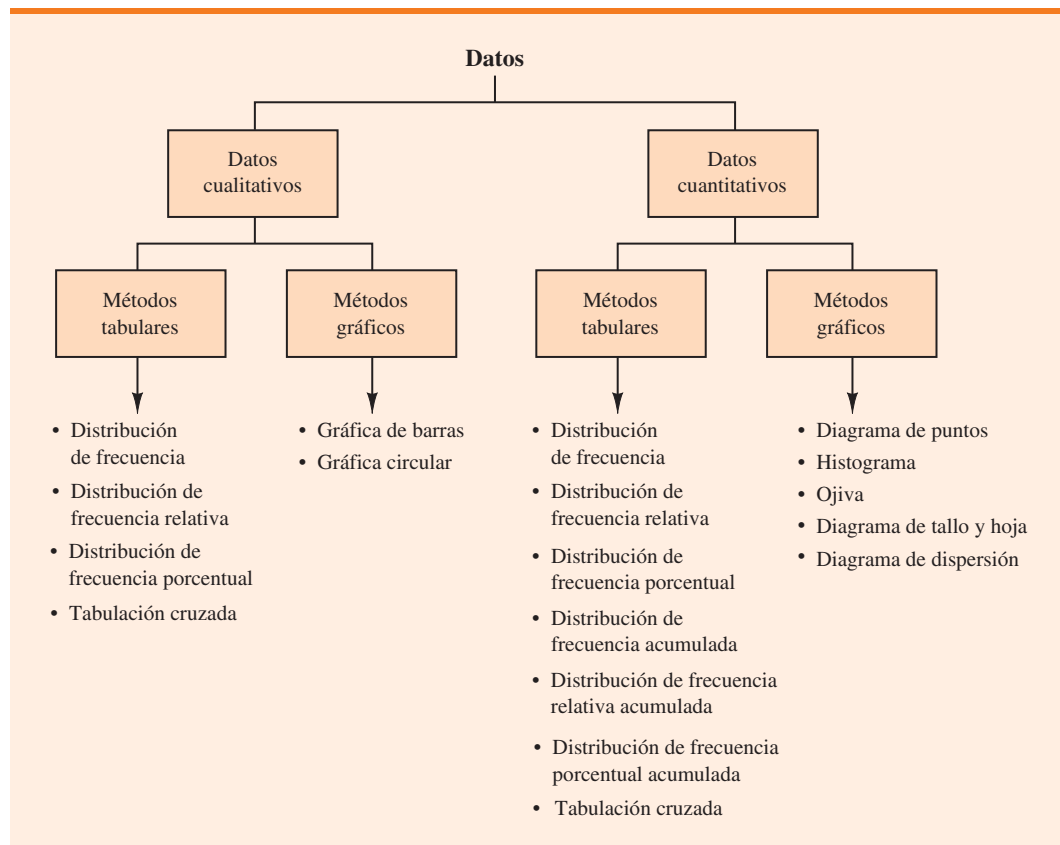
WEB **archivo**  
FuelData08

- c) Elabore una tabulación cruzada de los datos de tracción (filas) y las millas por galón en la ciudad (columnas). Use las clases 5–9, 10–14, 15–19, 20–24, 25–29, 30–34 y 35–39 para las millas por galón en la ciudad.
- d) Comente la relación entre la tracción y las millas por galón en la ciudad.
- e) Prepare una tabulación cruzada de los datos sobre el tipo de combustible (filas) y las millas por galón en la ciudad (columnas). Use las clases 5–9, 10–14, 15–19, 20–24, 25–29, 30–34 y 35–39 para las millas por galón en la ciudad.
- f) Comente la relación entre el tipo de combustible y las millas por galón en la ciudad.
38. Remítase al ejercicio 37 y a los datos en el archivo FuelData08.
- a) Elabore una tabulación cruzada de los datos sobre el desplazamiento (filas) y las millas por galón en autopista (columnas). Use las clases 1.0–2.9, 3.0–4.9 y 5.0–6.9 para el desplazamiento. Use las clases 15–19, 20–24, 25–29, 30–34 y 35–39 para las millas por galón en autopista.
- b) Comente la relación, si existe, entre el desplazamiento y las millas por galón en autopista.
- c) Elabore un diagrama de dispersión de los datos sobre el desplazamiento y las millas por galón en autopista. Use el eje vertical para las millas.
- d) ¿Qué indica el diagrama de dispersión elaborado en el inciso c) sobre la relación, si existe, entre el desplazamiento y las millas por galón en autopista?
- e) En la investigación de la relación entre el desplazamiento y las millas por galón en autopista usted elaboró un resumen tabular de los datos (tabulación cruzada) y un resumen gráfico (diagrama de dispersión). En este caso, ¿cuál método prefiere? Explique sus razones.

## Resumen

Con frecuencia es difícil interpretar directamente un conjunto de datos, incluso si es pequeño, en la forma en que se recolecta. Los métodos tabulares y gráficos proporcionan procedimientos para la organización y el resumen de los datos de modo que los patrones se inviertan y los datos se interpreten con más facilidad. Las distribuciones de frecuencia, las distribuciones de frecuencia relativa, las distribuciones de frecuencia porcentual, las gráficas de barras y las gráficas circulares se presentan como procedimientos tabulares y gráficos para el resumen de datos cualitativos. Las distribuciones de frecuencia, de frecuencia relativa y de frecuencia porcentual, así como los histogramas, las distribuciones de frecuencia acumulada, las distribuciones de frecuencia relativa acumulada, las distribuciones de frecuencia porcentual acumulada y las ojivas constituyen maneras de resumir datos cuantitativos. El diagrama de tallo y hoja proporciona una técnica de análisis explicativo de los datos que también se utiliza para resumir datos cuantitativos. La tabulación cruzada constituye un método tabular para resumir datos para dos variables. El diagrama de dispersión se plantea como un método gráfico para mostrar la relación entre dos variables cuantitativas. La figura 2.9 muestra los métodos tabulares y gráficos expuestos en este capítulo.

FIGURA 2.9 Métodos tabulares y gráficos para el resumen de datos



Con conjuntos de datos grandes, el software de computadora es fundamental para la elaboración de resúmenes tabulares y gráficos de los datos. En los apéndices del capítulo se muestra cómo se usan Minitab, Excel y StatTools para este propósito.

## Glosario

**Análisis de datos exploratorios** Métodos que utilizan aritmética sencilla y gráficas fáciles de trazar para resumir los datos de manera rápida.

**Datos cualitativos** Etiquetas o nombres que sirven para identificar categorías de elementos parecidos.

**Datos cuantitativos** Valores numéricos que indican cuánto o cuántos.

**Diagrama de dispersión** Presentación gráfica de la relación entre dos variables cuantitativas. Una variable se muestra en el eje horizontal y la otra en el eje vertical.

**Diagrama de puntos** Dispositivo gráfico que resume los datos según el número de puntos arriba de cada valor de datos en el eje horizontal.

**Diagrama de tallo y hoja** Técnica de análisis de datos exploratorios que clasifica de manera simultánea los órdenes de los datos cuantitativos y permite comprender la forma de la distribución.

**Distribución de frecuencia** Resumen tabular de los datos que muestra el número (frecuencia) de los valores de datos en cada una de varias clases que no se superponen.

**Distribución de frecuencia acumulada** Resumen tabular de datos cuantitativos que muestra el número de valores de datos que son menores o iguales que el límite de clase superior de cada clase.

**Distribución de frecuencia porcentual** Resumen tabular de los datos que muestra el porcentaje de valores de datos en cada una de varias clases que no se superponen.

**Distribución de frecuencia porcentual acumulada** Resumen tabular de los datos cuantitativos que muestra el porcentaje de los valores de datos que son menores o iguales que el límite de clase superior de cada clase.

**Distribución de frecuencia relativa** Resumen tabular de los datos que registra la fracción o proporción de los valores de datos en cada una de varias clases que no se superponen.

**Distribución de frecuencia relativa acumulada** Resumen tabular de los datos cuantitativos que muestra la fracción o proporción de los valores de datos que son menores o iguales al límite de clase superior de cada clase.

**Gráfica circular** Dispositivo gráfico para representar resúmenes de datos con base en la subdivisión de un círculo en sectores que corresponden a la frecuencia relativa de cada clase.

**Gráfica de barras** Dispositivo para representar datos cualitativos previamente resumidos en una distribución de frecuencia, distribución de frecuencia relativa o distribución de frecuencia porcentual.

**Histograma** Presentación gráfica de una distribución de frecuencia, distribución de frecuencia relativa o distribución de frecuencia porcentual de datos cuantitativos, elaborada mediante la colocación de los intervalos de clase en el eje horizontal y las frecuencias, frecuencias relativas o frecuencias porcentuales en el eje vertical.

**Línea de tendencia** Línea que proporciona una aproximación de la relación entre dos variables.

**Ojiva** Gráfica de una distribución acumulada.

**Paradoja de Simpson** Conclusiones obtenidas de dos o más tabulaciones cruzadas separadas que pueden revocarse cuando los datos se agregan en una sola tabulación cruzada.

**Punto medio de clase** Valor intermedio entre los límites de clase inferior y superior.

**Tabulación cruzada** Resumen tabular de los datos para dos variables. Las clases para una de las variables se representan por medio de filas; las clases para la otra variable se representan por medio de columnas.

## Fórmulas clave

### Frecuencia relativa

$$\frac{\text{Frecuencia de la clase}}{n} \quad (2.1)$$

### Ancho de clase aproximado

$$\frac{\text{Valor de datos mayor/valor de datos menor}}{\text{número de clases}} \quad (2.2)$$

## Ejercicios complementarios

39. El Instituto de Investigación de Educación Superior de la Universidad de California en Los Ángeles (UCLA) proporciona estadísticas sobre las asignaturas más populares entre los estudiantes de primer año que asisten a la universidad. Las cinco materias principales son arte y humanidades (A), administración de empresas (B), ingeniería (E), profesional (P) y ciencias sociales (S) (*The New York Times Almanac*, 2006). Una amplia variedad de otras asignaturas principales (O), que incluyen biología, física, ciencias de la computación y educación, se agrupan juntas. Las asignaturas principales seleccionadas por una muestra de 64 universitarios de primer año se presentan a continuación.

S	P	P	O	B	E	O	E	P	O	O	B	O	O	O	A
O	E	E	B	S	O	B	O	A	O	E	O	E	O	B	P
B	A	S	O	E	A	B	O	S	S	O	O	E	B	O	B
A	E	B	E	A	A	P	O	O	E	O	B	B	O	P	B

- Elabore una distribución de frecuencia y una distribución de frecuencia porcentual.
- Trace una gráfica de barras.

**WEB** archivo  
GMSales

- c) ¿Qué porcentaje de estudiantes de primer año seleccionó una de las cinco asignaturas principales más populares?
- d) ¿Cuál es la materia principal más popular para los universitarios de primer año? ¿Qué porcentaje de estudiantes la seleccionó?
40. General Motors tuvo un participación de 23% en la industria automotriz, con ventas que provienen de ocho divisiones: Buick, Cadillac, Chevrolet, GMC, Hummer, Pontiac, Saab y Saturn (*Forbes*, 22 de diciembre de 2008). El conjunto de datos de GMSales registra las ventas para una muestra de 200 vehículos de General Motors. La división por vehículo se proporciona para cada venta.
- a) Prepare la distribución de frecuencia y la distribución de frecuencia porcentual de las ventas por división para General Motors.
- b) Muestre una gráfica de barras de la distribución de frecuencia porcentual.
- c) ¿Cuál división de General Motors fue la empresa líder en ventas? ¿Cuál fue el porcentaje de ventas para esta división? ¿Fue la división más importante de General Motors? Explique por qué.
- d) Debido a la recesión actual, los altos precios de la gasolina y la disminución en las ventas de automóviles, General Motors enfrentó una bancarrota en 2009. Se anticiparon por ende el rescate financiero del gobierno y una reestructuración de la empresa. Las expectativas eran que ésta no podía seguir operando las ocho divisiones. Con base en el porcentaje de ventas, ¿cuál de las ocho divisiones parecería ser la mejor candidata a ser suspendida por General Motors? ¿Y cuáles divisiones parecerían ser las candidatas menos probables para suspensión?
41. El rendimiento de dividendos es el dividendo anual pagado por una empresa expresado como un porcentaje del precio de la acción ( $\text{dividendo/precio de la acción} \times 100$ ). El rendimiento de dividendos para las empresas del promedio industrial Dow Jones se muestra en la tabla 2.15 (*The Wall Street Journal*, 8 de junio de 2009).
- a) Elabore una distribución de frecuencia y una distribución de frecuencia porcentual.
- b) Prepare un histograma.
- c) Comente la forma de la distribución.
- d) ¿Qué indican los resúmenes tabulares y gráficos sobre los rendimientos de dividendos entre las empresas del promedio industrial Dow Jones?
- e) ¿Cuál empresa tiene el dividendo más alto producido? Si las acciones de ésta se venden actualmente a \$20 por acción y usted compra 500, ¿cuánto ingreso por dividendos generará esta inversión en un año?
42. Aproximadamente 1.5 estudiantes de secundaria y bachillerato presentan cada año el examen de aptitudes escolares (scholastic aptitude test, SAT), y casi 80% de los colegios y universidades sin políticas de admisión abiertas utilizan estas calificaciones en la toma de decisiones

**TABLA 2.15** Rendimiento de dividendos para las empresas del promedio industrial Dow Jones

Company	Dividend Yield %	Company	Dividend Yield %
3M	3.6	IBM	2.1
Alcoa	1.3	Intel	3.4
American Express	2.9	J.P. Morgan Chase	0.5
AT&T	6.6	Johnson & Johnson	3.6
Bank of America	0.4	Kraft Foods	4.4
Boeing	3.8	McDonald's	3.4
Caterpillar	4.7	Merck	5.5
Chevron	3.9	Microsoft	2.5
Cisco Systems	0.0	Pfizer	4.2
Coca-Cola	3.3	Procter & Gamble	3.4
DuPont	5.8	Travelers	3.0
ExxonMobil	2.4	United Technologies	2.9
General Electric	9.2	Verizon	6.3
Hewlett-Packard	0.9	Wal-Mart Stores	2.2
Home Depot	3.9	Walt Disney	1.5

**WEB** archivo  
DYield

de ingreso (College Board, marzo de 2009). La versión actual del SAT incluye tres partes: comprensión de lectura, matemáticas y redacción. Una calificación combinada perfecta para las tres partes es 2 400. Una muestra de calificaciones para el SAT combinado de tres partes es la siguiente.

**WEB** archivo  
NewSAT

1 665	1 525	1 355	1 645	1 780
1 275	2 135	1 280	1 060	1 585
1 650	1 560	1 150	1 485	1 990
1 590	1 880	1 420	1 755	1 375
1 475	1 680	1 440	1 260	1 730
1 490	1 560	940	1 390	1 175

- a) Elabore una distribución de frecuencia y un histograma. Comience con la primera clase en 800 y utilice un ancho de clase de 200.
- b) Comente la forma de la distribución.
- c) ¿Qué otras observaciones puede hacer acerca de las calificaciones del SAT con base en los resúmenes gráfico y tabular?
43. Los Acereros de Pittsburgh derrotaron a los Cardenales de Arizona 27 a 23 en el Super Bowl 43 del futbol americano. Con esta victoria, su sexto campeonato, los Acereros de Pittsburgh se convirtieron en el equipo con más triunfos en los 43 años de historia del evento (*Tampa Tribune*, 2 de febrero de 2009). El Super Bowl se ha celebrado en ocho estados diferentes de Estados Unidos: Arizona (AZ), California (CA), Florida (FL), Georgia (GA), Louisiana (LA), Michigan (MI), Minnesota (MN) y Texas (TX). Los datos de la tabla siguiente muestran el estado donde se celebraron los Super Bowls y el margen de puntos de la victoria para el equipo ganador.

**WEB** archivo  
SuperBowl

Super Bowl	State	Won By Points	Super Bowl	State	Won By Points	Super Bowl	State	Won By Points
1	CA	25	16	MI	5	31	LA	14
2	FL	19	17	CA	10	32	CA	7
3	FL	9	18	FL	19	33	FL	15
4	LA	16	19	CA	22	34	GA	7
5	FL	3	20	LA	36	35	FL	27
6	FL	21	21	CA	19	36	LA	3
7	CA	7	22	CA	32	37	CA	27
8	TX	17	23	FL	4	38	TX	3
9	LA	10	24	LA	45	39	FL	3
10	FL	4	25	FL	1	40	MI	11
11	CA	18	26	MN	13	41	FL	12
12	LA	17	27	CA	35	42	AZ	3
13	FL	4	28	GA	17	43	FL	4
14	CA	12	29	FL	23			
15	LA	17	30	AZ	10			

- a) Elabore una distribución de frecuencia y una gráfica de barra para el estado donde se celebró el Super Bowl.
- b) ¿A qué conclusiones llega a partir del resumen del inciso a)? ¿Cuál es el porcentaje de Super Bowls celebrados en los estados de Florida o California? ¿Qué porcentaje se celebró en los estados del norte o de clima frío?
- c) Muestre un diagrama de tallo y hoja para el margen de puntos de victoria para el equipo ganador. Elabore un histograma.
- d) ¿Qué conclusiones obtiene a partir de su resumen del inciso c)? ¿Qué porcentaje de Super Bowls han sido partidos cerrados con un margen de victoria menor que 5 puntos? ¿Qué porcentaje se ha ganado por 20 puntos o más?
- e) El partido de Super Bowl más cerrado tuvo lugar cuando los Gigantes de Nueva York derrotaron a los Bills de Búfalo. ¿Dónde se realizó este partido y cuál fue el margen de victoria del equipo ganador? El margen de puntos más grande en la historia del Super Bowl ocurrió cuando los 49's de San Francisco derrotaron a los Broncos de Denver. ¿En dónde se celebró este partido y cuál fue el margen de victoria del equipo ganador?

44. Los datos de la Oficina del Censo de Estados Unidos proporcionan la población por estado en millones de personas (*The World Almanac*, 2006).

**WEB** archivo  
Population

State	Population	State	Population	State	Population
Alabama	4.5	Louisiana	4.5	Ohio	11.5
Alaska	0.7	Maine	1.3	Oklahoma	3.5
Arizona	5.7	Maryland	5.6	Oregon	3.6
Arkansas	2.8	Massachusetts	6.4	Pennsylvania	12.4
California	35.9	Michigan	10.1	Rhode Island	1.1
Colorado	4.6	Minnesota	5.1	South Carolina	4.2
Connecticut	3.5	Mississippi	2.9	South Dakota	0.8
Delaware	0.8	Missouri	5.8	Tennessee	5.9
Florida	17.4	Montana	0.9	Texas	22.5
Georgia	8.8	Nebraska	1.7	Utah	2.4
Hawaii	1.3	Nevada	2.3	Vermont	0.6
Idaho	1.4	New Hampshire	1.3	Virginia	7.5
Illinois	12.7	New Jersey	8.7	Washington	6.2
Indiana	6.2	New Mexico	1.9	West Virginia	1.8
Iowa	3.0	New York	19.2	Wisconsin	5.5
Kansas	2.7	North Carolina	8.5	Wyoming	0.5
Kentucky	4.1	North Dakota	0.6		

- a) Elabore una distribución de frecuencia, una distribución de frecuencia porcentual y un histograma. Utilice un ancho de clase de 2.5 millones.
- b) Comente la tendencia en la distribución.
- c) ¿Qué observaciones puede hacer acerca de la población de los 50 estados?
45. *Drug Store News* (septiembre de 2002) proporcionó datos sobre las ventas farmacéuticas anuales para los minoristas de farmacias líderes en Estados Unidos. Los datos siguientes muestran las ventas anuales en millones de dólares.

Minorista	Ventas	Minorista	Ventas
Ahold USA	\$ 1700	Medicine Shoppe	\$ 1757
CVS	12700	Rite-Aid	8637
Eckerd	7739	Safeway	2150
Kmart	1863	Walgreens	11660
Kroger	3400	Wal-Mart	7250

- a) Elabore un diagrama de tallo y hoja.
- b) Identifique los niveles de ventas anuales para los minoristas de farmacia pequeños, medianos y grandes.
- c) ¿Cuáles son los dos minoristas más grandes?
46. Las temperaturas alta y baja diarias para 20 ciudades se listan a continuación (*USA Today*, 3 de marzo de 2006).

**WEB** archivo  
CityTemp

City	High	Low	City	High	Low
Albuquerque	66	39	Los Ángeles	60	46
Atlanta	61	35	Miami	84	65
Baltimore	42	26	Minneapolis	30	11
Charlotte	60	29	Nueva Orleans	68	50
Cincinnati	41	21	Oklahoma City	62	40
Dallas	62	47	Phoenix	77	50
Denver	60	31	Portland	54	38
Houston	70	54	St. Louis	45	27
Indianapolis	42	22	San Francisco	55	43
Las Vegas	65	43	Seattle	52	36

- a) Elabore un diagrama de tallo y hoja de las temperaturas altas.  
 b) Elabore un diagrama de tallo y hoja para las temperaturas bajas.  
 c) Compare los dos diagramas y comente sobre la diferencia entre las temperaturas altas y bajas.  
 d) Proporcione una distribución de frecuencia tanto para las temperaturas altas como para las bajas.
47. Consulte el conjunto de datos para las temperaturas altas y bajas de las 20 ciudades del ejercicio 46.  
 a) Elabore un diagrama de dispersión para mostrar la relación entre las dos variables: temperatura alta y temperatura baja.  
 b) Comente sobre la relación entre ambas temperaturas.
48. Una de las preguntas en una encuesta de *Financial Times/Harris Poll* fue: “¿Qué tanto está a favor o en contra de un impuesto mayor sobre las emisiones de carbono de los automóviles?” Las respuestas posibles fueron totalmente a favor, más a favor que en contra, más en contra que a favor y totalmente en contra. La tabulación cruzada siguiente muestra las respuestas obtenidas para 5 372 adultos encuestados en cuatro países de Europa y en Estados Unidos (sitio web de Harris Interactive, 27 de febrero de 2008).

Nivel de apoyo	País					Total
	Gran Bretaña	Italia	España	Alemania	Estados Unidos	
Totalmente a favor	337	334	510	222	214	1 617
Más a favor que en contra	370	408	355	411	327	1 871
Más en contra que a favor	250	188	155	267	275	1 135
Totalmente en contra	130	115	89	211	204	749
<b>Total</b>	1 087	1 045	1 109	1 111	1 020	5 372

- a) Elabore una distribución de frecuencia porcentual para la variable del nivel de apoyo. ¿Piensa usted que los resultados muestran apoyo a un impuesto más alto sobre la emisión de carbono de los automóviles?
- b) Elabore una distribución de frecuencia porcentual para la variable país.
- c) ¿El nivel de apoyo entre los adultos en los países europeos difiere del que se manifiesta en Estados Unidos? Explique por qué.
49. Western University sólo tiene una beca para jugadoras de softbol femenino para el próximo año. Las dos finalistas que la universidad está considerando son Allison Fealey y Emily Janson. El personal técnico ha concluido que la velocidad y las habilidades defensivas de las dos jugadoras son prácticamente idénticas, y que la decisión final se basará en la que tenga el mejor promedio de bateo. Las tabulaciones cruzadas del desempeño de bateo de cada jugadora en la secundaria y el bachillerato son las siguientes.

Resultado	Allison Fealey		Resultado	Emily Janson	
	Junior	Senior		Junior	Senior
Hit	15	75	Hit	70	35
Sin hit	25	175	Sin hit	130	85
<b>Total de bateos</b>	40	250	<b>Total de bateos</b>	200	120

El promedio de bateo se calcula al dividir el número de hits de un jugador por el número total de turnos al bate. Los promedios de bateo se representan como un número decimal con tres lugares después del punto decimal.

- a) Calcule el promedio de bateo de cada jugadora en la secundaria. Luego calcule el promedio de bateo en el bachillerato. Utilice este análisis para responder cuál jugadora debe obtener la beca. Explique sus razones.

- b) Combine o agregue los datos para la secundaria y el bachillerato en una tabulación cruzada de la manera siguiente.

Resultado	Jugadora	
	Fealey	Janson
Hit		
Sin hit		
Total de bateos		

Calcule el promedio de bateo de cada jugadora para los dos años combinados. Utilice este análisis para responder cuál de ellas debe recibir la beca. Explique sus razones.

- c) ¿Sus recomendaciones de los incisos a) y b) son congruentes? Justifique cualquier falta aparente de congruencia.
50. En una encuesta de edificios comerciales realizada por la Cincinnati Gas & Electric Company se preguntaba cuál era el combustible utilizado en la calefacción principal y en qué año se construyó el edificio. Una tabulación cruzada parcial de los hallazgos se presenta a continuación.

Año de construcción	Tipo de combustible				
	Electricidad	Gas natural	Petróleo	Propano	Otros
1973 o antes	40	183	12	5	7
1974-1979	24	26	2	2	0
1980-1986	37	38	1	0	6
1987-1991	48	70	2	0	1

- a) Complete la tabulación cruzada mostrando el total de las filas y el total de las columnas.
- b) Elabore las distribuciones de frecuencia del año de construcción y del tipo de combustible.
- c) Desarrolle una tabulación cruzada que muestre los porcentajes de columna.
- d) Elabore una tabulación cruzada que muestre los porcentajes de fila.
- e) Comente la relación entre el año de construcción y el tipo de combustible.
51. La tabla 2.16 incluye una porción de los datos contenidos en el archivo Fortune, y lista los datos sobre el capital de los accionistas, el valor de mercado y las utilidades para una muestra de 50 empresas Fortune 500.

**TABLA 2.16** Datos para una muestra de 50 empresas Fortune 500

Company	Stockholders' Equity (\$1000s)	Market Value (\$1000s)	Profit (\$1000s)
AGCO	982.1	372.1	60.6
AMP	2698.0	12017.6	2.0
Apple Computer	1642.0	4605.0	309.0
Baxter International	2839.0	21743.0	315.0
Bergen Brunswick	629.1	2787.5	3.1
Best Buy	557.7	10376.5	94.5
Charles Schwab	1429.0	35340.6	348.5
.	.	.	.
.	.	.	.
.	.	.	.
Walgreen	2849.0	30324.7	511.0
Westvaco	2246.4	2225.6	132.0
Whirlpool	2001.0	3729.4	325.0
Xerox	5544.0	35603.7	395.0

**WEB** archivo  
Fortune



- a) Elabore una tabulación cruzada para las variables del capital de los accionistas (Stockholders' Equity) y de las utilidades (Profit). Use las clases 0–200, 200–400, . . . , 1 000–1 200 para las utilidades, y las clases 0–1 200, 1 200–2 400, . . . , 4 800–6 000 para el capital de los accionistas.
  - b) Calcule los porcentajes de fila para la tabulación cruzada que elaboró para el inciso a).
  - c) ¿Qué relación observa, si hay alguna, entre las utilidades y el capital de los accionistas?
52. Remítase al conjunto de datos de la tabla 2.16.
- a) Elabore una tabulación cruzada para las variables valor de mercado (Market Value) y utilidades (Profit).
  - b) Calcule los porcentajes de fila para su tabulación cruzada del inciso a).
  - c) Comente sobre cualquier relación entre las variables.
53. Consulte el conjunto de datos de la tabla 2.16.
- a) Trace un diagrama de dispersión que muestre la relación entre las variables utilidades y capital de los accionistas.
  - b) Comente acerca de cualquier relación entre las variables.
54. Consulte el conjunto de datos de la tabla 2.16.
- a) Elabore un diagrama de dispersión que muestre la relación entre las variables valor de mercado y capital de los accionistas.
  - b) Comente sobre cualquier relación entre las variables.

## Caso a resolver 1 Pelican Stores

Pelican Stores, una división de National Clothing, es una cadena de tiendas de ropa para mujer que opera en todo Estados Unidos. La cadena lanzó recientemente una promoción en la que se enviaron cupones de descuento a los clientes de otras tiendas de National Clothing. Los datos recabados de una muestra de 100 transacciones de tarjetas de crédito en Pelican Stores durante un día, mientras la promoción estuvo vigente, se encuentran en el archivo llamado *PelicanStores*. La tabla 2.17 muestra una parte del conjunto de datos.

El método de pago Proprietary Card se refiere a los cargos realizados con una tarjeta de National Clothing. A los clientes que efectuaron una compra utilizando un cupón de descuento se les llama *clientes de promoción* y a los que hicieron una compra pero no usaron un cupón de descuento se les llama *clientes habituales*. Dado que los cupones promocionales no se enviaron a los clientes regulares de Pelican Stores, la gerencia considera las ventas realizadas a personas que presentaron los cupones como ventas que de lo contrario no se hubieran efectuado. Por supuesto, Pelican también espera que los clientes de promoción sigan comprando en sus tiendas.

**TABLA 2.17** Datos para una muestra de 100 compras con tarjeta de crédito en Pelican Stores

Customer	Type of Customer	Items	Net Sales	Method of Payment	Gender	Marital Status	Age
1	Regular	1	39.50	Discover	Male	Married	32
2	Promotional	1	102.40	Proprietary Card	Female	Married	36
3	Regular	1	22.50	Proprietary Card	Female	Married	32
4	Promotional	5	100.40	Proprietary Card	Female	Married	28
5	Regular	2	54.00	MasterCard	Female	Married	34
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
96	Regular	1	39.50	MasterCard	Female	Married	44
97	Promotional	9	253.00	Proprietary Card	Female	Married	30
98	Promotional	10	287.59	Proprietary Card	Female	Married	52
99	Promotional	2	47.60	Proprietary Card	Female	Married	30
100	Promotional	1	28.44	Proprietary Card	Female	Married	44

**WEB** archivo  
PelicanStores

La mayoría de las variables mostradas en la tabla 2.17 se explican por sí mismas, pero dos de ellas requieren una aclaración.

Artículos (Items)	Número total de productos adquiridos
Ventas netas (Net Sales)	Monto total (\$) cargado a la tarjeta de crédito

A la gerencia de Pelican le gustaría usar estos datos muestrales para enterarse de su base de clientes y evaluar la promoción que consiste en los cupones de descuento.

### Informe gerencial

Use los métodos tabular y gráfico de la estadística descriptiva para ayudar a la gerencia a elaborar un perfil de los clientes y evaluar la campaña promocional. Como mínimo, su informe debe incluir lo siguiente:

1. Distribución de frecuencia porcentual para las variables clave.
2. Una gráfica de barras o circular que muestre el número de compras del cliente atribuible al método de pago.
3. Una tabulación cruzada del tipo de cliente (regular o promocional) contra las ventas netas. Comente cualquier similitud o diferencia que se presente.
4. Un diagrama de dispersión para explorar la relación entre las ventas netas (Net sales) y la edad (Age) de los clientes.

## Caso a resolver 2 Industria del cine

La industria fílmica estadounidense es un negocio muy competitivo. Más de 50 estudios producen un total de 300 a 400 películas nuevas cada año, y el éxito financiero de cada una varía considerablemente. Las ventas brutas del fin de semana de estreno (en millones de dólares), las ventas brutas totales (Total Gross Sales) (en millones de dólares), el número de salas (Number of Theaters) donde se exhibe la película y el número de semanas en que ésta permaneció entre las primeras 60 (Weeks in the Top 60) en ventas brutas son variables comunes utilizadas para medir el éxito de una cinta. Los datos recabados de una muestra de 100 películas producidas en 2005 se incluyen en el archivo llamado *Movies*. La tabla 2.18 muestra los datos de las primeras 10 películas de este archivo.

### Informe gerencial

Use los métodos tabulares y gráficos de la estadística descriptiva para conocer cómo estas variables contribuyen al éxito de una película. Incluya en su informe los puntos que se indican en la siguiente página.

**TABLA 2.18** Datos del desempeño de 10 películas

Motion Picture	Opening Gross Sales (\$millions)	Total Gross Sales (\$millions)	Number of Theaters	Weeks in Top 60
<i>Coach Carter</i>	29.17	67.25	2574	16
<i>Ladies in Lavender</i>	0.15	6.65	119	22
<i>Batman Begins</i>	48.75	205.28	3858	18
<i>Unleashed</i>	10.90	24.47	1962	8
<i>Pretty Persuasion</i>	0.06	0.23	24	4
<i>Fever Pitch</i>	12.40	42.01	3275	14
<i>Harry Potter and the Goblet of Fire</i>	102.69	287.18	3858	13
<i>Monster-in-Law</i>	23.11	82.89	3424	16
<i>White Noise</i>	24.11	55.85	2279	7
<i>Mr. and Mrs. Smith</i>	50.34	186.22	3451	21

1. Los resúmenes tabulares y gráficos para cada una de las cuatro variables junto con un análisis de cada resumen que proporcionen información sobre la industria del cine.
2. Un diagrama de dispersión para explorar la relación entre las ventas brutas totales y las ventas brutas de estreno. Explíquelo.
3. Un diagrama de dispersión para explorar la relación entre las ventas brutas totales y el número de cines. Coméntelo.
4. Un diagrama de dispersión para explorar la relación entre las ventas brutas totales y el número de semanas en que la película permaneció entre las primeras 60. Coméntelo.

## Apéndice 2.1 **Uso de Minitab para presentaciones tabulares y gráficas**

Minitab ofrece amplias capacidades para elaborar resúmenes tabulares y gráficos de los datos. En este apéndice se muestra cómo se usa para elaborar varios resúmenes gráficos y tabulares de una tabulación cruzada. Los métodos gráficos presentados incluyen el diagrama de puntos, el histograma, el diagrama de tallo y hoja, el diagrama de dispersión y la tabulación cruzada.

### Diagrama de puntos

**WEB** archivo

Audit

Para esta demostración se utilizan los datos de duración de la auditoría de la tabla 2.4. Los datos están en la columna C1 de una hoja de trabajo de Minitab. Los pasos siguientes generarán un diagrama de puntos.

- Paso 1.** Seleccione el menú **Graph** y elija **Dotplot**.
- Paso 2.** Seleccione **One Y, Simple** y haga clic en **OK**.
- Paso 3.** Cuando el cuadro de diálogo Dotplot-One Y, Simple aparezca:  
Introduzca C1 en el cuadro **Graph Variables**.  
Haga clic en **OK**.

### Histograma

**WEB** archivo

Audit

En esta sección se muestra cómo elaborar un histograma con las frecuencias sobre el eje vertical usando los datos de duración de la auditoría de la tabla 2.4. Los datos se encuentran en la columna C1 de la hoja de trabajo de Minitab. Los pasos siguientes generarán un histograma para la duración de las auditorías.

- Paso 1.** Seleccione el menú **Graph**.
- Paso 2.** Elija **Histogram**.
- Paso 3.** Seleccione **Simple** y haga clic en **OK**.
- Paso 4.** Cuando el cuadro de diálogo Histogram-Simple se abra:  
Introduzca C1 en el cuadro **Graph Variables**.  
Haga clic en **OK**.
- Paso 5.** Cuando aparezca el histograma:  
Coloque el puntero del mouse sobre cualquiera de las barras.  
Haga doble clic.
- Paso 6.** Cuando el cuadro de diálogo Edit Bars (editar barras) aparezca:  
Haga clic en la ficha **Binning**.  
Seleccione **Cutpoint** para el tipo de intervalo (Interval Type).  
Seleccione **Midpoint/Cutpoint positions** para la definición del intervalo (Interval Definition).  
Introduzca 10:35/5 en el cuadro **Midpoint/Cutpoint positions**.<sup>\*</sup>  
Haga clic en **OK**.

<sup>\*</sup> La entrada 10:35/5 indica que 10 es el valor inicial para el histograma, 35 es el valor final para el histograma y 5 es el ancho de clase.

Observe que Minitab también proporciona la opción de aumentar o disminuir el eje  $x$  de modo que los valores numéricos aparezcan en los puntos medios de los rectángulos del histograma. Si desea activar esta opción, modifique el paso 6 para incluir **Select Midpoint** para el tipo de intervalo e introduzca 12:32/5 en el cuadro **Midpoint/Cutpoint positions**. Estos pasos producen el mismo histograma con los puntos medios de los rectángulos rotulados como 12, 17, 22, 27 y 32.

## Diagrama de tallo y hoja

### WEB archivo

ApTest

Los datos de la prueba de aptitudes de la tabla 2.8 se usan para mostrar la elaboración de un diagrama de tallo y hoja. Los datos se encuentran en la columna C1 de la hoja de trabajo de Minitab. Los pasos siguientes generarán el diagrama de tallo y hoja extendido que se muestra en la sección 2.3.

**Paso 1.** Seleccione el menú **Graph**.

**Paso 2.** Elija **Stem-and-Leaf**.

**Paso 3.** Cuando el cuadro de diálogo Stem-and-Leaf aparezca:  
Introduzca C1 en el cuadro **Graph Variables**.  
Haga clic en **OK**.

## Diagrama de dispersión

### WEB archivo

Stereo

Los datos de la tienda de estéreos y equipos de sonido de la tabla 2.12 se usan para mostrar la elaboración de un diagrama de dispersión. Las semanas están numeradas del 1 al 10 en la columna C1; los datos del número de comerciales están en la columna C2, y los datos de las ventas en la columna C3 de la hoja de trabajo de Minitab. Los pasos siguientes generan el diagrama de dispersión mostrado en la figura 2.7.

**Paso 1.** Seleccione el menú **Graph**.

**Paso 2.** Elija **Scatterplot**.

**Paso 3.** Seleccione **Simple** y haga clic en **OK**.

**Paso 4.** Cuando el cuadro de diálogo Scatterplot-Simple se abra:  
Introduzca C3 bajo **Y variables** y C2 bajo **X variables**.  
Haga clic en **OK**.

## Tabulación cruzada

### WEB archivo

Restaurant

Para esta demostración se utilizan los datos de la revisión del restaurante Zagat's, parte de los cuales se presentan en la tabla 2.9. Los restaurantes se numeran del 1 al 300 en la columna C1 de la hoja de trabajo de Minitab. Las calificaciones de calidad están en la columna C2 y los precios de los alimentos en la columna C3.

Minitab sólo puede crear una tabulación cruzada para variables cualitativas, y el precio de la comida es una variable cuantitativa. Así que primero necesitamos codificar los datos de la segunda variable al especificar la clase a la cual pertenece el precio de cada comida. Los pasos siguientes codificarán los datos del precio de los alimentos para crear cuatro clases en la columna C4: \$10–19, \$20–29, \$30–39 y \$40–49.

**Paso 1.** Seleccione el menú **Data**.

**Paso 2.** Elija **Code**.

**Paso 3.** Elija **Numeric to Text**.

**Paso 4.** Cuando el cuadro de diálogo Code-Numeric to Text aparezca:

Introduzca C3 en el cuadro **Code data from columns**.

Dé enter a C4 en el cuadro **Store coded data in columns**.

Introduzca 10:19 en el primer cuadro **Original values** y \$10-19 en el cuadro **New** adyacente.

Introduzca 20:29 en el segundo cuadro **Original values** y \$20-29 en el cuadro **New** adyacente.

Introduzca 30:39 en el tercer cuadro **Original values** y \$30-39 en el cuadro **New** adyacente.

Introduzca 40:49 en el cuarto cuadro **Original values** y \$40-49 en el cuadro **New** adyacente.

Haga clic en **OK**.

En la columna C4 aparecerá la categoría de precio asociada con cada precio de comida de la columna C3. Ahora podemos elaborar una tabulación cruzada para la calificación de la calidad y las categorías de precio de los alimentos usando los datos de las columnas C2 y C4. Los pasos siguientes crearán una tabulación cruzada que contiene la misma información que la mostrada en la tabla 2.10.

**Paso 1.** Seleccione el menú **Stat**.

**Paso 2.** Elija **Tables**.

**Paso 3.** Elija **Cross Tabulation and Chi-Square**.

**Paso 4.** Cuando el cuadro de diálogo Cross Tabulation and Chi-Square se abra:

Introduzca C2 en el cuadro **For rows** y C4 en el cuadro **For columns**.

Seleccione **Counts** bajo **Display**.

Haga clic en **OK**.

## Apéndice 2.2 Uso de Excel para presentaciones tabulares y gráficas

Excel ofrece muchas capacidades para elaborar resúmenes de datos tabulares y gráficos. En este apéndice se muestra cómo se usa este programa para elaborar una distribución de frecuencia, una gráfica de barras, una gráfica circular, un histograma, un diagrama de dispersión y una tabulación cruzada. Veremos el uso de tres de las herramientas más poderosas de Excel para el análisis de datos: herramientas de graficación y los informes de tablas dinámicas y gráficas dinámicas.

### Distribución de frecuencia y gráfica de barras de datos cualitativos

En esta sección se muestra cómo se usa Excel para elaborar una distribución de frecuencia y una gráfica de barras de datos cualitativos. Ilustramos cómo se utilizan los datos sobre las compras de bebidas refrescantes de la tabla 2.1.

**Distribución de frecuencia** Comenzamos por mostrar el uso de la función COUNTIF para elaborar una distribución de frecuencia de los datos de la tabla 2.1. Vuelva a observar la figura 2.10 a medida que se describan los pasos requeridos. La hoja de trabajo de la fórmula (donde aparecen las funciones y las fórmulas usadas) se coloca en segundo plano, y la hoja de trabajo de valores (donde se muestran los resultados obtenidos mediante las funciones y las fórmulas) aparece en primer plano.

La etiqueta “Brand Purchased” y los datos para la compra de 50 bebidas refrescantes se encuentran en las celdas A1:A51. También se introdujeron las etiquetas “Soft Drink” y “Frequency” en las celdas C1:D1. Los cinco nombres de bebidas refrescantes se introducen en las celdas C2:C6. La función COUNTIF de Excel se usa entonces para contar el número de veces que aparece cada bebida en las celdas A2:A51. Siga estos pasos:

**Paso 1.** Seleccione la celda D2.

**Paso 2.** Introduzca =COUNTIF(\$A\$2:\$A\$51,C2).

**Paso 3.** Copie la celda D2 a las celdas D3:D6.

La hoja de trabajo de la fórmula de la figura 2.10 muestra las fórmulas de las celdas que se insertan al aplicar estos pasos. La hoja de trabajo de valores registra los valores calculados por las fórmulas de las celdas y presenta la misma distribución de frecuencia que se elaboró en la tabla 2.2.

**FIGURA 2.10** Distribución de frecuencia de la compra de bebidas refrescantes elaborada mediante la función COUNTIF de Excel

	A	B	C	D	E
1	<b>Brand Purchased</b>		<b>Soft Drink</b>	<b>Frequency</b>	
2	Coke Classic		Coke Classic	=COUNTIF(\$A\$2:\$A\$51,C2)	
3	Diet Coke		Diet Coke	=COUNTIF(\$A\$2:\$A\$51,C3)	
4	Pepsi		Dr. Pepper	=COUNTIF(\$A\$2:\$A\$51,C4)	
5	Diet Coke		Pepsi	=COUNTIF(\$A\$2:\$A\$51,C5)	
6	Coke Classic		Sprite	=COUNTIF(\$A\$2:\$A\$51,C6)	
7	Coke Classic				
8	Dr. Pepper				
9	Diet Coke				
10	Pepsi				
45	Pepsi				
46	Pepsi				
47	Pepsi				
48	Coke Classic				
49	Dr. Pepper				
50	Pepsi				
51	Sprite				
52					

	A	B	C	D	E
1	<b>Brand Purchased</b>		<b>Soft Drink</b>	<b>Frequency</b>	
2	Coke Classic		Coke Classic	19	
3	Diet Coke		Diet Coke	8	
4	Pepsi		Dr. Pepper	5	
5	Diet Coke		Pepsi	13	
6	Coke Classic		Sprite	5	
7	Coke Classic				
8	Dr. Pepper				
9	Diet Coke				
10	Pepsi				
45	Pepsi				
46	Pepsi				
47	Pepsi				
48	Coke Classic				
49	Dr. Pepper				
50	Pepsi				
51	Sprite				
52					

Nota. Las filas 11-44 están ocultas.

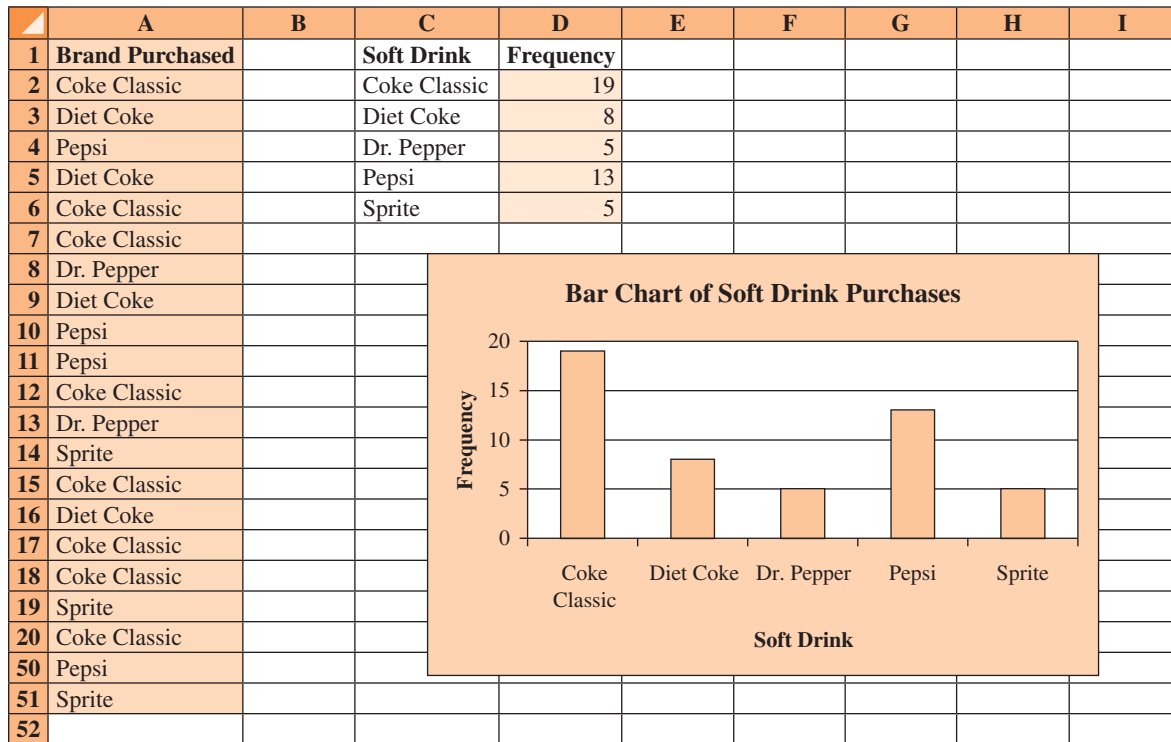
### WEB archivo

SoftDrink

**Gráfica de barras** Aquí se explica cómo se utilizan las herramientas de graficación de Excel para elaborar una gráfica de barras de los datos de las bebidas refrescantes. Consulte la distribución de frecuencia mostrada en la hoja de trabajo de valores de la figura 2.10. La gráfica de barras que se diseñará es una extensión de esta hoja de trabajo. La gráfica de barras y la hoja de trabajo elaboradas se presentan en la figura 2.11. Los pasos para realizarlas son los siguientes.

- Paso 1.** Seleccione las celdas C2:D6.
- Paso 2.** Haga clic en la ficha **Insert** de la cinta de opciones.
- Paso 3.** En el grupo **Charts**, haga clic en **Column**.
- Paso 4.** Cuando aparezca la lista de subtipos de gráficas de columna:
  - Vaya a la sección **2-D Column**.
  - Haga clic en la gráfica del extremo izquierdo, **Clustered Column**.
- Paso 5.** En el grupo **Chart Layouts**, haga clic en el botón **More** (la flecha que apunta hacia abajo con una línea sobre ella) para ver todas las opciones.
- Paso 6.** Elija el **Layout 9**.
- Paso 7.** Seleccione **Chart Title** y reemplace el nombre de la gráfica con **BarChart of Soft Drink Purchases**.
- Paso 8.** Elija **Horizontal (Category) Axis Title** y replácelo con **Soft Drink**.
- Paso 9.** Seleccione **Vertical (Value) Axis Title** y sustitúyalo con **Frequency**.
- Paso 10.** Haga clic con el botón secundario en **Series 1 Legend Entry**.
  - Haga clic en **Delete**.
- Paso 11.** Haga clic con el botón secundario en el eje vertical.
  - Haga clic en **Format Axis**.

**FIGURA 2.11** Gráfica de barras de la compra de bebidas refrescantes elaborada con las herramientas de graficación de Excel



**Paso 12.** Cuando el cuadro de diálogo Format Axis se abra:  
 Vaya a la sección **Axis Options**.  
 Seleccione **Fixed** para **Major Unit** e introduzca 5.0 en el cuadro correspondiente.  
 Haga clic en **Close**.

La gráfica de barras resultante se muestra en la figura 2.11.\* Excel puede producir una gráfica circular para los datos de la compra de bebidas refrescantes de una manera parecida. La principal diferencia radica en que en el paso 3 haría clic en **Pie** en el grupo **Charts**. Existen varios estilos de gráficas circulares.

### Distribución de frecuencia e histograma para datos cuantitativos

*En una sección posterior de este apéndice se describe cómo usar PivotTable Report de Excel para elaborar una tabulación cruzada.*

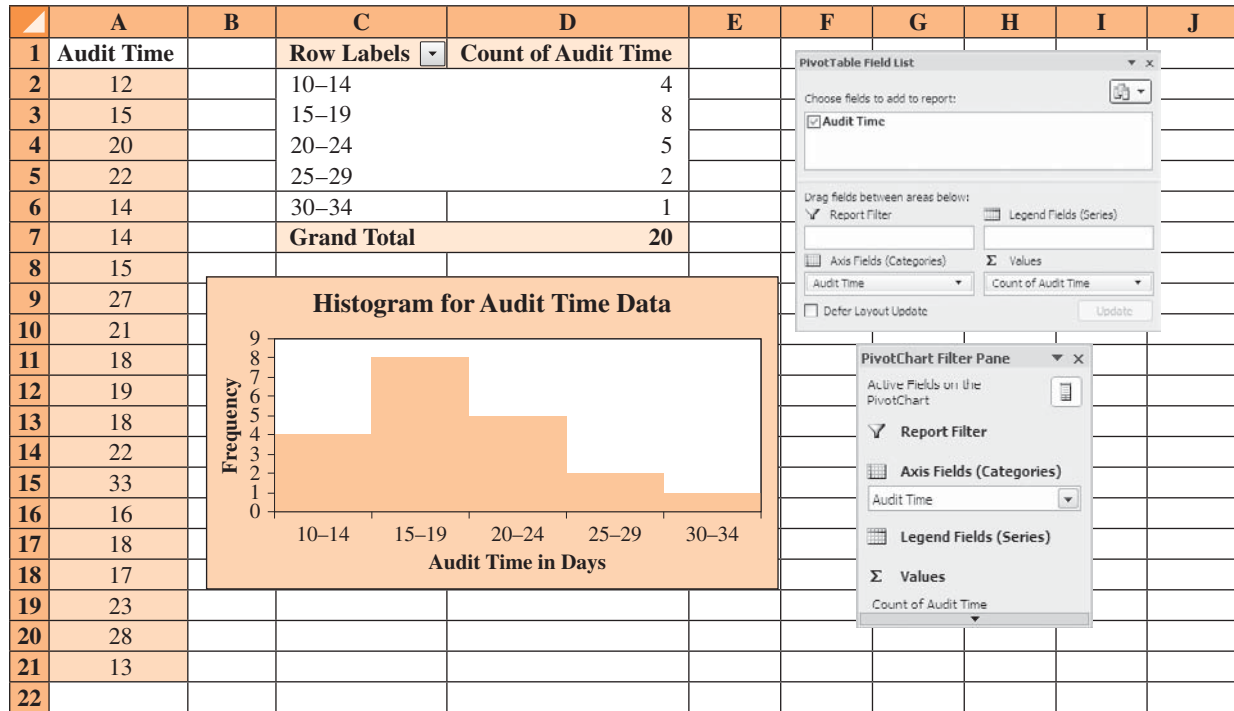
El informe de tabla dinámica de Excel (PivotTable Report) es una herramienta interactiva que permite resumir los datos de manera rápida en una variedad de maneras, que incluyen la elaboración de una distribución de frecuencia para datos cuantitativos. Una vez que se crea una distribución de frecuencia utilizando el informe de tabla dinámica, entonces se usan las herramientas de graficación de Excel para preparar el histograma correspondiente. No obstante, el informe de gráfico dinámico de Excel permite elaborar de forma simultánea una distribución de frecuencia y un histograma. Ilustraremos este procedimiento usando los datos de duración de la auditoría de la tabla 2.4. La etiqueta "Audit Time" y los 20 valores de la duración de la auditoría se introdujeron en las celdas A1:A21 de la hoja de cálculo de Excel. Los pasos listados en la siguiente página describen cómo usar el informe de gráfico dinámico de Excel para preparar una distribución de frecuencia y un histograma de dichos datos. Vuelva a observar la figura 2.12 cuando se describan los pasos correspondientes.

**WEB** archivo

**Audit**

\* El tamaño de la gráfica de barras de la figura 2.11 puede modificarse. Cambiar el tamaño de una gráfica en Excel no es difícil. Primero seleccione la gráfica. Los controladores de tamaño aparecerán en el borde de ésta. Haga clic en los controladores y arrástrelos para modificar el tamaño de la figura como lo desee.

**FIGURA 2.12** Uso del informe de gráfico dinámico de Excel para elaborar una distribución de frecuencia y un histograma de los datos de duración de la auditoría



- Paso 1.** Haga clic en la ficha **Insert** de la cinta de opciones.
- Paso 2.** En el grupo **Tables**, haga clic en la palabra **PivotTable**.
- Paso 3.** Elija **PivotChart** de las opciones que aparecen.
- Paso 4.** Cuando el cuadro de diálogo **Create PivotTable with PivotChart** aparezca:
  - Seleccione **Select a table or range**.
  - Introduzca A1:A21 en el cuadro **Table/Range**.
  - Elija **Existing Worksheet** como la ubicación para la PivotTable y PivotChart.
  - Introduzca C1 en el cuadro **Location**.
  - Haga clic en **OK**.
- Paso 5.** En **PivotTable Field List**, vaya a **Choose Fields to add to report**.
  - Arrastre el campo **Audit Time** al área **Axis Fields (Categories)**.
  - Arrastre el campo **Audit Time** al área **Values**.
- Paso 6.** Haga clic en **Sum of Audit Time** en la sección **Values**.
- Paso 7.** Dé clic en **Value Field Settings** de la lista de opciones que aparece.
- Paso 8.** Cuando el cuadro de diálogo **Value Field Settings** se abra:
  - Bajo **Summarize value field by**, elija **Count**.
  - Haga clic en **OK**.
- Paso 9.** Cierre la **PivotTable Field List**.
- Paso 10.** Haga clic con el botón secundario del mouse en la celda C2 en el informe de PivotTable o en cualquier otra celda que contenga una duración de la auditoría.
- Paso 11.** Elija **Group** de la lista de opciones que aparezca.
- Paso 12.** Cuando se abra el cuadro de diálogo **Grouping**.
  - Introduzca 10 en el cuadro **Starting at**.



Introduzca 34 en el cuadro **Ending at**.  
 Introduzca 5 en el cuadro **By**.  
 Haga clic en **OK** (aparecerá un gráfico dinámico).

- Paso 13.** Haga clic dentro del **PivotChart** resultante.
- Paso 14.** Haga clic en la ficha **Design** en la cinta de opciones.
- Paso 15.** En el grupo **Chart Layouts**, haga clic en el botón **More** (la flecha que apunta hacia abajo con una línea sobre ella) para mostrar todas las opciones.
- Paso 16.** Elija **Layout 8**.
- Paso 17.** Seleccione el **Chart Title** y replácelo con **Histogram for Audit Time Data**.
- Paso 18.** Seleccione **Horizontal (Category) Axis Title** y sustitúyalo con **Audit Time in Days**.
- Paso 19.** Elija el título **Vertical (Value) Axis Title** y replácelo con **Frequency**.

La figura 2.12 muestra los informes de tabla dinámica y gráfico dinámico resultantes. Observamos que el informe de tabla dinámica proporciona la distribución de frecuencia de los datos de duración de la auditoría y el informe de gráfico dinámico proporciona el histograma correspondiente. Si lo desea, puede cambiar las etiquetas de cualquier celda en la distribución de frecuencia al seleccionar la celda y teclear la etiqueta nueva.

### Tabulación cruzada

El informe de tabla dinámica de Excel es una manera excelente de resumir los datos para dos o más variables de forma simultánea. Se explicará el uso de este informe al mostrar cómo elaborar una tabulación cruzada de las calificaciones de calidad y los precios de la comida de la muestra de 300 restaurantes de Los Ángeles. Se usarán los datos del archivo llamado *Restaurant*; las etiquetas “Restaurant”, “Quality Rating” (calificación de la calidad) y “Meal Price (\$)” (precio de la comida) se introdujeron en las celdas A1:C1 de la hoja de cálculo como se aprecia en la figura 2.13. Los datos de cada uno de los restaurantes de la muestra se introdujeron en las celdas B2:C301.

**FIGURA 2.13** Hoja de cálculo de Excel que contiene datos de los restaurantes

**WEB** archivo  
 Restaurant

	A	B	C	D
1	Restaurant	Quality Rating	Meal Price (\$)	
2	1	Good	18	
3	2	Very Good	22	
4	3	Good	28	
5	4	Excellent	38	
6	5	Very Good	33	
7	6	Good	28	
8	7	Very Good	19	
9	8	Very Good	11	
10	9	Very Good	23	
11	10	Good	13	
292	291	Very Good	23	
293	292	Very Good	24	
294	293	Excellent	45	
295	294	Good	14	
296	295	Good	18	
297	296	Good	17	
298	297	Good	16	
299	298	Good	15	
300	299	Very Good	38	
301	300	Very Good	31	
302				

**Nota.** Las filas 12-291 están ocultas.

Si se desea utilizar el informe de tabla dinámica para elaborar una tabulación cruzada, se requiere realizar tres tareas: mostrar la lista de campos de la tabla dinámica inicial y el informe de tabla dinámica; establecer la lista de campos de la tabla dinámica, y finalizar el informe de tabla dinámica. Estas tareas se describen enseguida.

**Mostrar la lista de campos de la tabla dinámica inicial y el informe de tabla dinámica.** Se requieren tres pasos para mostrar la lista de campos inicial y el informe de tabla dinámica.

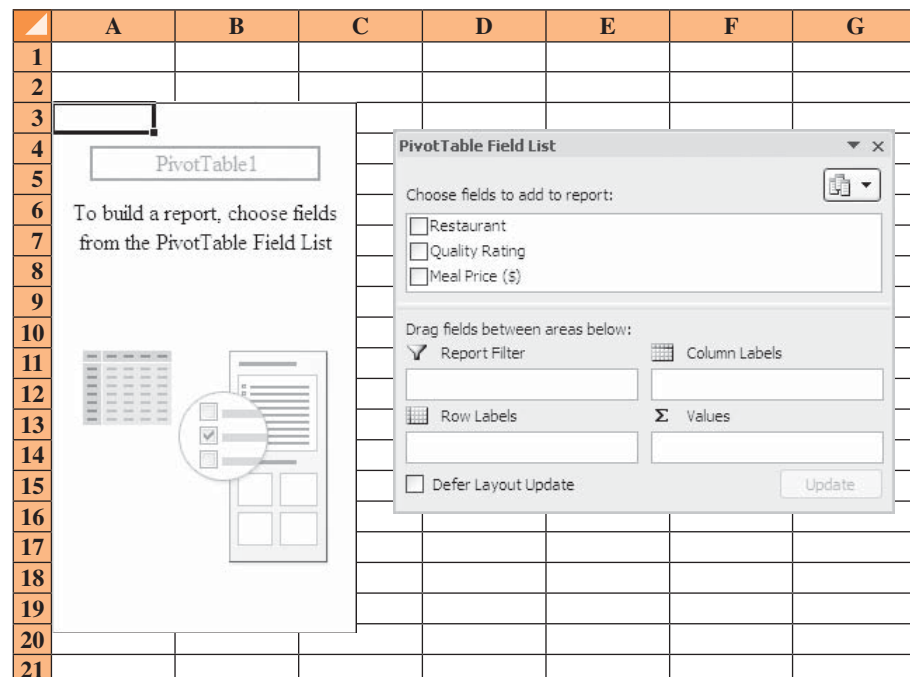
- Paso 1.** Haga clic en la ficha **Insert** en la cinta de opciones.
- Paso 2.** En el grupo **Tables**, haga clic en el icono sobre la palabra PivotTable.
- Paso 3.** Cuando el cuadro de diálogo **Create PivotTable** aparezca:
  - Elija **Select a Table or Range**.
  - Introduzca A1:C301 en el cuadro **Table/Range**.
  - Elija **New Worksheet** como la ubicación para PivotTable Report.
  - Haga clic en **OK**.

La lista de campo inicial de tabla dinámica y el informe de tabla dinámica se muestran en la figura 2.14.

**Configuración de la lista de campos inicial de tabla dinámica.** Excel considera cada una de las tres columnas de la figura 2.13 [etiquetadas como Restaurant, Quality Rating y Meal Price (\$)] como un campo. Los campos se eligen para representar filas, columnas o valores en el cuerpo del informe de tabla dinámica. Los pasos siguientes muestran cómo utilizar la lista de campos de tabla dinámica de Excel para asignar el campo Quality Rating a las filas, el campo Meal Price (\$) a las columnas y el campo Restaurant al cuerpo del informe de la tabla dinámica.

- Paso 1.** En **PivotTable Field List**, vaya a **Choose Fields to add to report**.
  - Arrastre el campo **Quality Rating** a la sección **Row Labels**.
  - Arrastre el campo **Meal Price (\$)** a la sección **Column Labels**.
  - Arrastre el campo **Restaurant** a la sección **Values**.

**FIGURA 2.14** Lista de campo inicial e informe de campo de PivotTable para los datos del restaurante



- Paso 2.** Haga clic con el botón secundario en **Sum of Restaurant** en la sección **Values**.
- Paso 3.** Haga clic en la opción **Value Field Settings** de la lista de opciones que se despliega.
- Paso 4.** Cuando el cuadro de diálogo **Value Field Settings** se abra:  
 Bajo **Summarize value field by**, elija **Count**.  
 Haga clic en **OK**.

La figura 2.15 muestra la lista de campos de tabla dinámica completada y una porción de la hoja de trabajo de tabla dinámica tal como aparece.

**Finalizar el informe de tabla dinámica.** Para completar el informe de tabla dinámica se necesita agrupar las columnas que representan los precios de la comida y colocar las etiquetas de fila para la calificación de la calidad en el orden apropiado. Considere los pasos siguientes para hacerlo.

- Paso 1.** Haga clic con el botón secundario en la celda B4 o en cualquier otra que contenga precios de comida.
- Paso 2.** Elija **Group** en la lista de opciones que aparece.
- Paso 3.** Cuando el cuadro de diálogo **Grouping** se abra:  
 Introduzca 10 en el cuadro **Starting at**.  
 Introduzca 49 en el cuadro **Ending at**.  
 Introduzca 10 en el cuadro **By**.  
 Haga clic en **OK**.
- Paso 4.** Haga clic con el botón secundario en **Excellent** en la celda A5.
- Paso 5.** Elija **Move** y haga clic en **Move “Excellent” to End**.

El informe de tabla dinámica final se muestra en la figura 2.16. Observe que contiene la misma información que la tabulación cruzada de la tabla 2.10.

### Diagrama de dispersión

Las herramientas de graficación de Excel se utilizaron para elaborar un diagrama de dispersión y una línea de tendencia de los datos de la tienda de estéreos y equipos de sonido presentados

**FIGURA 2.15** Lista de campos de la tabla dinámica completada y una porción del informe de tabla dinámica de los datos del restaurante (las columnas H:AK están ocultas)

	A	B	C	D	E	F	G	AL	AM	AN	AO
1											
2											
3	<b>Count of Restaurant</b>	Column Labels									
4	Row Labels	10	11	12	13	14	15	47	48	Grand Total	
5	Excellent				1			2	2	66	
6	Good	6	4	3	3	2	4			84	
7	Very Good	1	4	3	5	6	1		1	150	
8	<b>Grand Total</b>	<b>7</b>	<b>8</b>	<b>6</b>	<b>9</b>	<b>8</b>	<b>5</b>	<b>2</b>	<b>3</b>	<b>300</b>	
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											

**PivotTable Field List**

Choose fields to add to report:

- Restaurant
- Quality Rating
- Meal Price (\$)

Drag fields between areas below:

Report Filter

Column Labels: Meal Price (\$)

Row Labels: Quality Rating

Values: Count of Restaurant

Defer Layout Update

Update

FIGURA 2.16 Informe de tabla dinámica final de los datos del restaurante

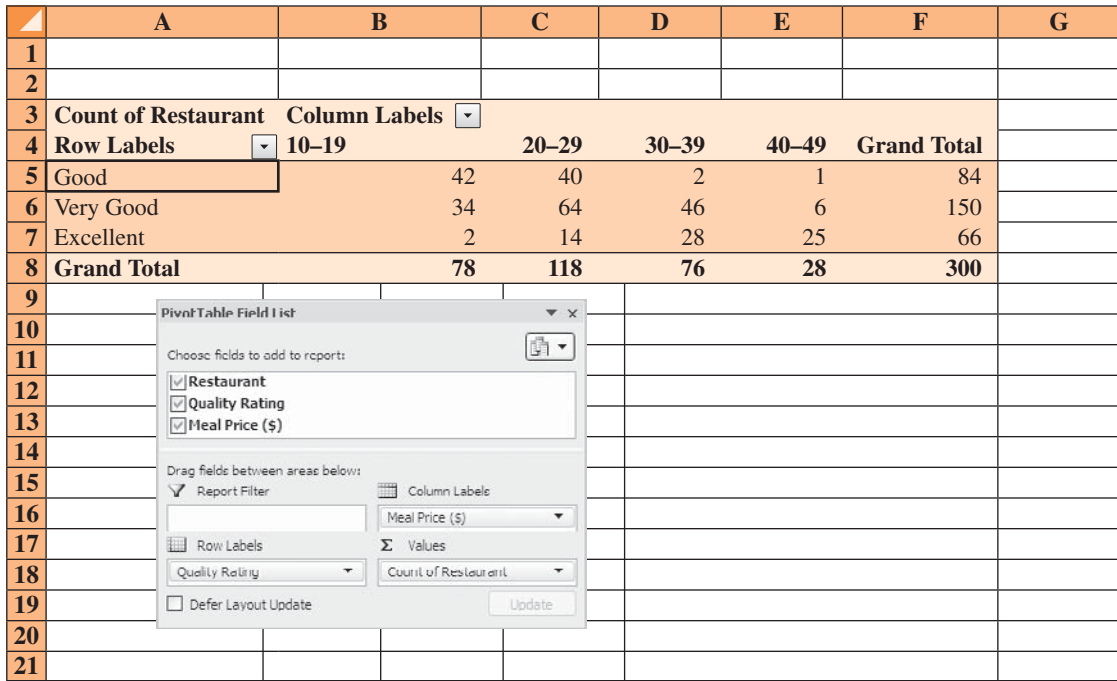
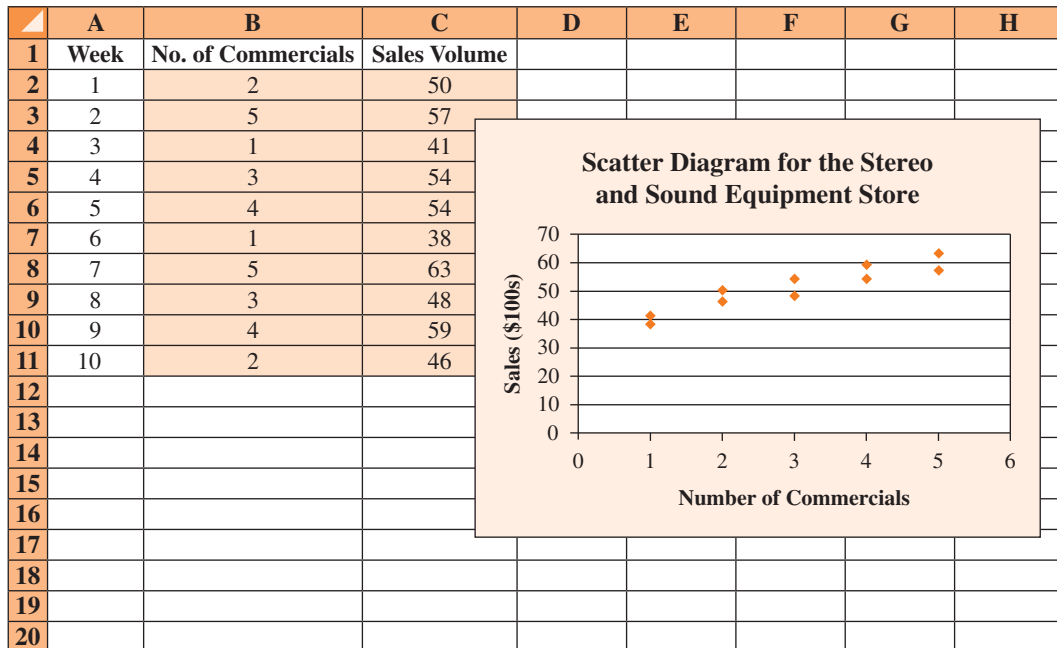
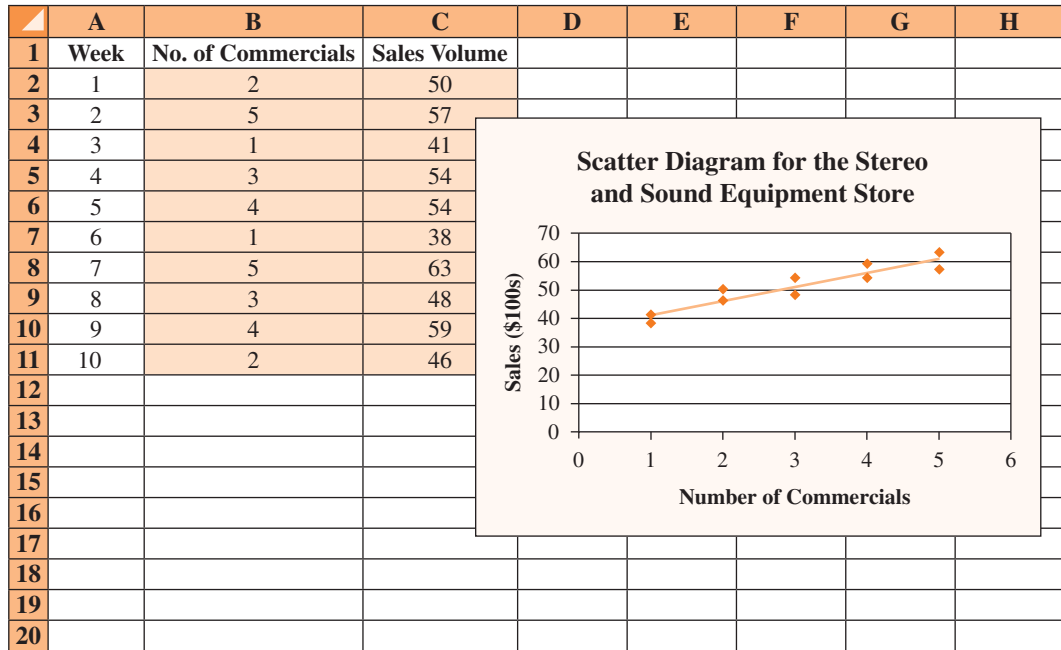


FIGURA 2.17 Diagrama de dispersión de la tienda de estéreos y equipos de sonido usando las herramientas de graficación de excel



**FIGURA 2.18** Diagrama de dispersión y línea de tendencia de la tienda de estéreo y equipos de sonido usando las herramientas de graficación de Excel



en la tabla 2.12. Vuelva a observar las figuras 2.17 y 2.18 a medida que se describan los pasos correspondientes. Usaremos los datos del archivo llamado *Stereo*; las etiquetas Week, No. of Commercials y Sales Volume se introdujeron en las celdas A1:C1 de la hoja de cálculo. Los datos de cada una de las 10 semanas se introdujeron en las celdas B2:C11. Los pasos siguientes describen cómo usar las herramientas de graficación de Excel para producir un diagrama de dispersión de los datos.

- Paso 1.** Seleccione las celdas B2:C11.
- Paso 2.** Haga clic en la ficha **Insert** en la cinta de opciones.
- Paso 3.** En el grupo **Charts**, haga clic en **Scatter**.
- Paso 4.** Cuando la lista de subtipos de diagramas de dispersión se abra, haga clic en **Scatter with only Markers** (la tabla de la esquina superior izquierda).
- Paso 5.** En el grupo **Chart Layouts**, haga clic en **Layout 1**.
- Paso 6.** Elija el **Chart Title** y replácelo con **Scatter Diagram for the Stereo and Sound Equipment Store**.
- Paso 7.** Seleccione **Horizontal (Value) Axis Title** y replácelo con **Number of Commercials**.
- Paso 8.** Seleccione **Vertical (Value) Axis Title** y replácelo con **Sales (\$100s)**.
- Paso 9.** Haga clic con el botón secundario en **Series 1 Legend Entry** y haga clic en **Delete**.

La hoja de cálculo de la figura 2.17 muestra el diagrama de dispersión producido por Excel. Los pasos siguientes describen cómo añadir una línea de tendencia.

- Paso 1.** Coloque el puntero del *mouse* sobre cualquier punto de datos en el diagrama de dispersión y haga clic con el botón secundario para mostrar una lista de opciones.
- Paso 2.** Elija **Add Trendline**.
- Paso 3.** Cuando el cuadro de diálogo **Format Trendline** se abra:
  - Seleccione **Trendline Options**.
  - Elija **Linear** de la lista **Trend/Regression Type**.
  - Haga clic en **Close**.

La hoja de cálculo de la figura 2.18 muestra el diagrama de dispersión con la línea de tendencia agregada.

### Apéndice 2.3

## Uso de StatTools para presentaciones tabulares y gráficas

En este apéndice se explica cómo se utiliza StatTools para elaborar un histograma y un diagrama de dispersión.

### Histograma

Recurrimos a los datos de duración de la auditoría en la tabla 2.4 para la explicación. Comience usando Data Set Manager para crear un conjunto de datos StatTools para esos datos por medio del procedimiento descrito en el apéndice del capítulo 1. Los pasos siguientes generarán un histograma.

**WEB** archivo  
Audit

- Paso 1.** Haga clic en la ficha **StatTools** en la cinta de opciones.
- Paso 2.** En **Analyses Group**, haga clic en **Summary Graphs**.
- Paso 3.** Elija la opción **Histogram**.
- Paso 4.** Cuando el cuadro de diálogo StatTools–Histogram se abra:
  - En la sección **Variables**, seleccione **Audit Time**.
  - En la sección **Options**:
    - Introduzca 5 en el cuadro **Number of Bins**.
    - Introduzca 9.5 en el cuadro **Histogram Minimum**.
    - Introduzca 34.5 en el cuadro **Histogram Maximum**.
    - Elija **Categorical** en el cuadro **X-Axis**.
    - Elija **Frequency** en el cuadro **Y-Axis**.
  - Haga clic en **OK**.

Aparecerá un histograma para los datos de duración de la auditoría parecido al de la figura 2.12. La única diferencia es que el histograma elaborado usando StatTools muestra los puntos medios de clase en el eje horizontal.

### Diagrama de dispersión

**WEB** archivo  
Stereo

Para mostrar la elaboración de un diagrama de dispersión con StatTools se utilizan los datos de los estéreos y los equipos de sonido de la tabla 2.12. Comience con Data Set Manager para crear un conjunto de datos StatTools de estos datos mediante el procedimiento descrito en el apéndice del capítulo 1. Los pasos siguientes generarán un diagrama de dispersión.

- Paso 1.** Haga clic en la ficha **StatTools** en la cinta de opciones.
- Paso 2.** En **Analyses Group**, haga clic en **Summary Graphs**.
- Paso 3.** Elija la opción **Scatterplot**.
- Paso 4.** Cuando aparezca el cuadro de diálogo StatTools–Scatterplot:
  - En la sección **Variables**,
    - En la columna etiquetada **X**, seleccione **No. of Commercials**.
    - En la columna etiquetada **Y**, elija **Sales Volume**.
  - Haga clic en **OK**.

Aparecerá un diagrama de dispersión parecido al de la figura 2.17.