



CAPÍTULO 2

Estadística descriptiva: presentaciones tabulares y gráficas

CONTENIDO

LA ESTADÍSTICA EN LA PRÁCTICA: LA EMPRESA COLGATE-PALMOLIVE

2.1 RESUMEN DE DATOS CUALITATIVOS

Distribución de frecuencia
Distribuciones de frecuencia relativa y de frecuencia porcentual
Gráficas de barra y gráficas de pastel

2.2 RESUMEN DE DATOS CUANTITATIVOS

Distribución de frecuencia
Distribuciones de frecuencia relativa y de frecuencia porcentual

Gráficas de puntos

Histograma
Distribuciones acumuladas
Ojiva

2.3 ANÁLISIS EXPLORATORIO DE DATOS: EL DIAGRAMA DE TALLO Y HOJAS

2.4 TABULACIONES CRUZADAS Y DIAGRAMAS DE DISPERSIÓN

Tabulación cruzada
Paradoja de Simpson
Diagrama de dispersión y línea de tendencia



LA ESTADÍSTICA *(en)* LA PRÁCTICA

LA EMPRESA COLGATE-PALMOLIVE*

NUEVA YORK, NUEVA YORK

La empresa Colgate-Palmolive empezó en la Ciudad de Nueva York en 1806 como una pequeña tienda de jabones y velas. Hoy, Colgate-Palmolive emplea más de 4000 personas que trabajan en 200 países y territorios del mundo. Aunque es más conocida por sus marcas Colgate, Palmolive, Ajax y Fab, la empresa comercializa los productos Mennen, Hill's Science Diet y Hill's Prescription Diet.

La empresa Colgate-Palmolive aplica la estadística en su programa de aseguramiento de la calidad en los detergentes caseros para la ropa. Le interesa la satisfacción del cliente con la cantidad de detergente en los paquetes. Todos los paquetes de cierto tamaño se llenan con la misma cantidad de detergente en peso, aunque el volumen del detergente varía de acuerdo con la densidad del polvo detergente. Por ejemplo, si la densidad del detergente es alta, se necesita una cantidad menor de detergente para tener el peso señalado en el paquete. El resultado es que cuando el cliente abre el paquete le parece que no ha sido bien llenado.

Para controlar el problema del peso del polvo de detergente, se han establecido límites en el nivel aceptable de la densidad del polvo. Con periodicidad se toman muestras estadísticas y se mide la densidad de la muestra de polvo. Los resúmenes de los datos se les proporcionan a los operarios para que de ser necesario lleven a cabo acciones correctivas, de manera que la densidad se mantenga dentro de las especificaciones de calidad establecidas.

En la tabla y figura adjuntas se presentan una distribución de frecuencia y un histograma obtenidos con 150 muestras tomadas en una semana. Densidades mayores a 0.40 son inaceptablemente altas. De acuerdo con la distribución de frecuencia y al histograma la operación satisface los lineamientos de calidad ya que todas las densidades son menores o iguales a 0.40. A la vista de estos resúmenes estadísticos los directivos estarán satisfechos con la calidad del proceso de producción de detergente.

En este capítulo se estudiarán métodos tabulares y gráficos de la estadística descriptiva como distribuciones de frecuencia, gráficas de barras, histogramas, diagramas de tallo y hoja, tabulaciones cruzadas y otros. El objeto de



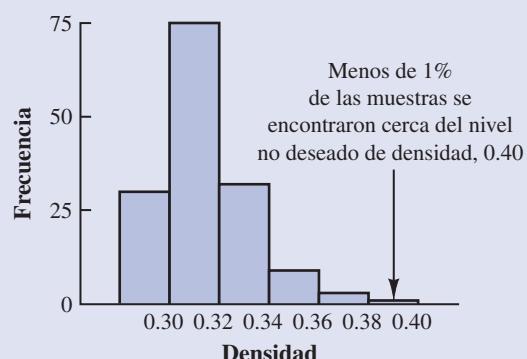
Los resúmenes estadísticos ayudan a mantener la calidad de estos productos de Colgate-Palmolive
© Joe Higgins/South Western.

estos métodos es resumir los datos de manera que sean entendibles e interpretables con facilidad.

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. Fawle, director de aseguramiento de la calidad de la empresa Colgate-Palmolive por proporcionarles este artículo para *La estadística en la práctica*.

Como se indicó en el capítulo 1, los datos se clasifican en cualitativos o cuantitativos. Los **datos cualitativos** emplean etiquetas o nombres para determinar categorías de elementos iguales. Los **datos cuantitativos** son números que indican cuánto o cuántos.

En este capítulo se presentan los métodos tabulares y gráficos empleados para datos cualitativos y cuantitativos. Los resúmenes gráficos o tabulares de datos se encuentran en reportes anuales, en artículos en los periódicos y en estudios de investigación. Todo mundo se encuentra con este tipo de presentaciones. Por tanto, es útil saber cómo se hacen y se interpretan. Se empezará con los métodos tabulares y gráficos para resumir datos que se refieren a una sola variable. En la última sección se introducen los métodos para resumir datos cuando lo que interesa es la relación entre dos variables.

Los paquetes modernos de software para estadística proporcionan muchas posibilidades para resumir datos y elaborar presentaciones gráficas. Minitab y Excel son dos paquetes muy empleados. En los apéndices de este capítulo se muestran algunas de sus posibilidades.

2.1

Resumen de datos cualitativos

Distribución de frecuencia

Conviene iniciar el estudio acerca del uso de los métodos tabulares y gráficos para resumir datos cualitativos 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 las diferentes clases disyuntas (que no se sobreponen).

Con el ejemplo siguiente se muestra la elaboración e interpretación de una distribución de frecuencia de datos cualitativos. Cinco refrescos muy conocidos son Coca cola clásica (Coke Classic), Coca cola de dieta (Diet Coke), Dr. Pepper, Pepsi y Sprite. Suponga que los datos de la tabla 2.1 muestran los refrescos que fueron comprados en una muestra de 50 ventas de refresco.

TABLA 2.1 DATOS DE UNA MUESTRA DE 50 VENTAS DE REFRESCO

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 LAS VENTAS DE REFRESCO

Refresco	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 aparece cada refresco en la tabla 2.1. La Coca cola clásica (Coke Classic) aparece 19 veces, la Coca cola de dieta (Diet Coke) 8 veces, Dr. Pepper 5 veces, Pepsi 13 veces y Sprite 5 veces. Esto queda resumido en la distribución de frecuencia de la tabla 2.2.

Esta distribución de frecuencia proporciona un resumen de cómo se distribuyeron las 50 ventas entre los cinco refrescos. El resumen aporta más claridad que los datos originales de la tabla 2.1. Al observar esta distribución de frecuencia, es claro que Coca cola clásica es el refresco que más se vende, Pepsi el segundo, Coca cola de dieta el tercero y Sprite y Dr. Pepper están empatados en el cuarto lugar. La distribución de frecuencia resume la información sobre la popularidad de los cinco refrescos.

Distribuciones de frecuencia relativa y de frecuencia porcentual

En una distribución de frecuencia se aprecia el número (frecuencia) de los elementos de cada una de las diversas clases disyuntas. Sin embargo, con frecuencia lo que interesa es la proporción o porcentaje de elementos en cada clase. La *frecuencia relativa* de una clase es igual a la parte o proporción de los elementos que pertenecen a cada clase. En un conjunto de datos, en el que hay 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.

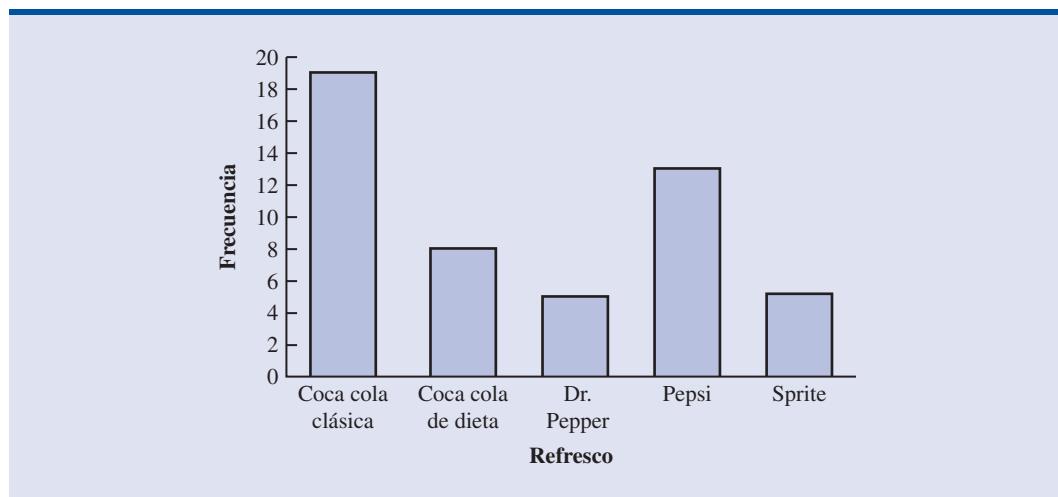
Una **distribución de frecuencia relativa** da un resumen tabular de datos en el que se muestra la frecuencia relativa de cada clase. Una **distribución de frecuencia porcentual** da la frecuencia porcentual de los datos de cada clase. En la tabla 2.3 se presenta una distribución de frecuencia relativa y una distribución de frecuencia porcentual de los datos de los refrescos. En esta tabla se observa que la frecuencia relativa de la Coca cola clásica es $19/50 = 0.38$, la de la Coca cola de dieta es $8/50 = 0.16$, etc. En la distribución de frecuencia porcentual, se muestra que 38% de las ventas fueron de Coca cola clásica, 16% de Coca cola de dieta, etc. También resulta que $38\% + 26\% + 16\% = 80\%$ de las ventas fueron de los tres refrescos que más se venden.

Gráficas de barra y gráficas de pastel

Una **gráfica de barras** o un diagrama de barras, es una gráfica para representar los datos cualitativos de una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. En uno de los ejes de la gráfica (por lo general en el horizontal), se especifican las etiquetas empleadas para las clases (categorías). Para el otro eje de la gráfica (el vertical) se usa una escala para

TABLA 2.3 DISTRIBUCIONES DE FRECUENCIA RELATIVA Y FRECUENCIA PORCENTUAL DE LAS VENTAS DE REFRESCOS

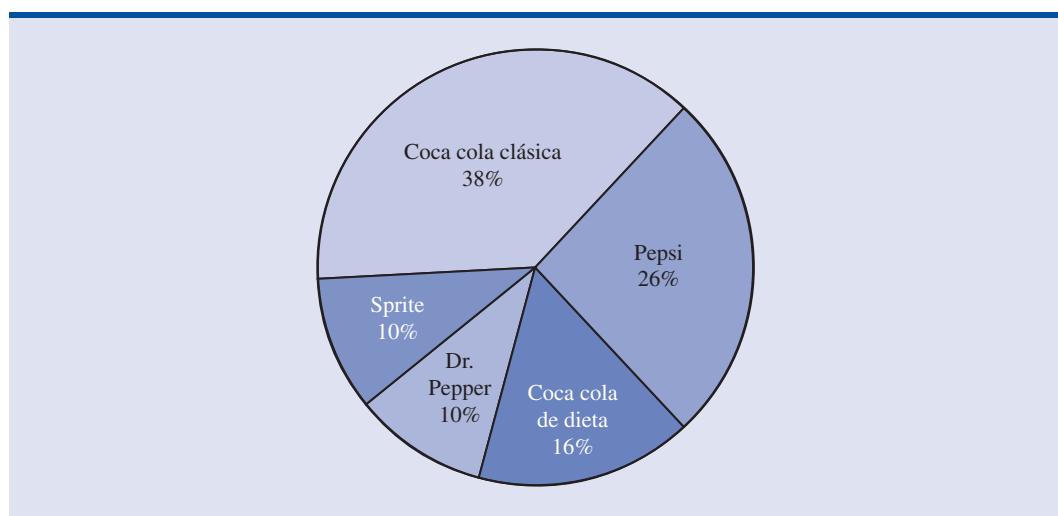
Refresco	Frecuencia relativa	Frecuencia porcentual
Coke Classic	0.38	38
Diet Coke	0.16	16
Dr. Pepper	0.10	10
Pepsi	0.26	26
Sprite	0.10	10
Total	1.00	100

FIGURA 2.1 GRÁFICA DE BARRAS PARA LAS VENTAS DE REFRESCOS

En el control de calidad, las gráficas de barras se usan para identificar las principales causas de problemas. Las graficas se acomodan en orden de alturas descendentes de izquierda a derecha colocando primero la causa de frecuencia más común en primer lugar. A esta gráfica de barras se le llama diagrama de Pareto en honor a su inventor Wilfredo Pareto, un economista italiano.

frecuencia, frecuencia relativa o frecuencia porcentual. Después, empleando un ancho de barra fijo, se dibuja sobre cada etiqueta de las clases una barra que se extiende hasta la frecuencia, frecuencia relativa o frecuencia porcentual de la clase. Cuando se tienen datos cualitativos, las barras deben estar separadas para hacer énfasis en que cada clase está separada. En la figura 2.1 se muestra una gráfica de barras correspondiente a la distribución de frecuencia de las 50 ventas de refrescos. Advierta cómo en esta representación gráfica se observa que Coca cola clásica, Pepsi y Coca cola de dieta son los refrescos preferidos.

La **gráfica de pastel** proporciona otra gráfica para presentar distribuciones de frecuencia relativa y de frecuencia porcentual de datos cualitativos. Para elaborar una gráfica de pastel, primero se dibuja un círculo que representa todos los datos. Después se usa la frecuencia relativa para subdividir el círculo en sectores, o partes, que corresponden a la frecuencia relativa de cada clase. Por ejemplo, como un círculo tiene 360 grados y Coca cola clásica presenta una frecuencia relativa de 0.38, el sector de la gráfica de pastel correspondiente a Coca cola clásica resultará de $0.38(360) = 136.8$ grados. El sector del pastel para Coca cola de dieta constará de

FIGURA 2.2 GRÁFICA DE PASTEL PARA LAS VENTAS DE REFRESCOS

$0.16(360) = 57.6$ grados. Mediante cálculos semejantes para las demás clases se obtiene la gráfica de pastel de la figura 2.2. Los números que aparecen en cada sector pueden ser frecuencia, frecuencia relativa o frecuencia porcentual.

NOTAS Y COMENTARIOS

1. A menudo el número de clases en una distribución de frecuencia es el mismo que el número de categorías encontradas en los datos, como en los datos de las ventas de refresco en esta sección. Los datos comprenden cinco refrescos y para cada uno se definió una clase en la distribución de frecuencia. Si los datos incluyeran todos los refrescos se requerirían muchas categorías, la mayor parte de las cuales sólo tendrían muy pocas ventas. La mayoría de los profesionistas de la estadística aconsejan que las clases con frecuencia pequeña, se agrupen en una sola clase a la que se le llama “otros”. Cualquier clase con 5% o menos se trata de esta manera.
2. La suma de las frecuencias en una distribución de frecuencia es siempre igual al número de observaciones. La suma de las frecuencias relativas en una distribución de frecuencia relativa es siempre igual a 1.00, y la suma de los porcentajes en una distribución de frecuencia porcentual es siempre igual a 100.

Ejercicios

Métodos

1. Como respuesta a una pregunta hay tres alternativas: A, B y C. En una muestra de 120 respuestas, 60 fueron A, 24 B y 36 C. Dé las distribuciones de frecuencia y de frecuencia relativa.
2. Se da una distribución de frecuencia relativa.

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 es 200. ¿Cuál es la frecuencia de la clase D?
 - c. Muestre la distribución de frecuencia.
 - d. Dé la distribución de frecuencia porcentual.
3. Un cuestionario proporciona como respuestas 58 Sí, 42 No y 20 ninguna opinión.
 - a. En la construcción de una gráfica de pastel, ¿cuántos grados le corresponderán del pastel a la respuesta Sí?
 - b. ¿Cuántos grados le corresponderán del pastel a la respuesta No?
 - c. Construya una gráfica de pastel.
 - d. Construya una gráfica de barras.

Aplicaciones

4. Los cuatro programas con horario estelar de televisión son *CSI*, *ER*, *Everybody Loves Raymond* y *Friends* (*Nielsen Media Research*, 11 de enero de 2004). A continuación se presentan los datos sobre las preferencias de los 50 televidentes de una muestra.

CSI	Friends	CSI	CSI	CSI
CSI	CSI	Raymond	ER	ER
Friends	CSI	ER	Friends	CSI
ER	ER	Friends	CSI	Raymond
CSI	Friends	CSI	CSI	Friends
ER	ER	ER	Friends	Raymond
CSI	Friends	Friends	CSI	Raymond
Friends	Friends	Raymond	Friends	CSI
Raymond	Friends	ER	Friends	CSI
CSI	ER	CSI	Friends	ER

- a. ¿Estos datos son cualitativos o cuantitativos?
- b. Proporcione las distribuciones de frecuencia y de frecuencia relativa.
- c. Construya una gráfica de barras y una gráfica de pastel.
- d. De acuerdo con la muestra, ¿qué programa de televisión tiene la mayor audiencia? ¿Cuál es el segundo?
5. Los cinco apellidos más comunes en Estados Unidos, en orden alfabético son, Brown, Davis, Johnson, Jones, Smith y Williams (*The World Almanac, 2006*). Suponga que en una muestra de 50 personas con uno de estos apellidos se obtienen 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 estos datos construyendo:

- a. Distribuciones de frecuencia relativa y porcentual.
- b. Una gráfica de barras.
- c. Una gráfica de pastel.
- d. De acuerdo con estos datos, ¿cuáles son los tres apellidos más comunes?
6. El índice de audiencia de televisión de Nielsen Media Research mide el porcentaje de personas que tienen televisión y que están viendo un determinado programa. El programa de televisión con el mayor índice de audiencia en la historia de la televisión (en Estados Unidos) fue *M*A*S*H Last Episode Special* transmitido el 28 de febrero de 1983. El índice de audiencia de 60.2 indicó que 60.2% de todas las personas que tenían televisión estaban viendo este programa. Nielsen Media Research publicó la lista de los 50 programas de televisión con los mayores índices de audiencia en la historia de la televisión (*The New York Times Almanac, 2006*). Los datos siguientes presentan las cadenas de televisión que produjeron estos 50 programas con mayor índice de audiencia.

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. Con estos datos construya una distribución de frecuencia, una de frecuencia porcentual y una gráfica de barras.

- b. ¿Cuál o cuáles cadenas de televisión han presentado los programas de mayor índice de audiencia? Compare los desempeños de ABC, CBS y NBC.

7. Un restaurante de Florida emplea cuestionarios en los que pide a sus clientes que evalúen el servicio, la calidad de los alimentos, los cocteles, los precios y la atmósfera del restaurante. Cada uno de estos puntos se evalúa con una escala de óptimo (O), muy bueno (V), bueno (G), regular (A) y malo (P). Emplee la estadística descriptiva para resumir los datos siguientes respecto a la calidad de los alimentos. ¿Qué piensa acerca de la evaluación de la calidad de los alimentos de este 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. A continuación se muestran datos de 55 miembros de un equipo de béisbol. Cada observación indica la posición principal que juegan los miembros del equipo: *pitcher* (P), *catcher* (H), primera base (1), segunda base (2), tercera base (3), shortstop (S), left fiel (L), center field (C) y right field (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. Para resumir estos datos use una distribución de frecuencia y otra de frecuencia relativa.
 b. ¿Cuál es la posición que ocupan más miembros del equipo?
 c. ¿Cuál es la posición que ocupan menos miembros del equipo?
 d. ¿Qué posición de campo (L, R, C) es la que juegan más miembros del equipo?
 e. Compare las posiciones L, R, y C con las posiciones 1, 2, 3 y S.
9. Cerca del 60% de las empresas pequeñas y medianas son empresas familiares. En un estudio de TEC International se preguntaba al gerente general (CEO, por sus siglas en inglés) cómo había llegado a ese cargo (*The Wall Street Journal*, 16 de diciembre de 2003). Las respuestas fueron que el CEO heredó el negocio, que el CEO formó la empresa o que el CEO estaba contratado por con la empresa. En una muestra de 26 CEOs de empresas familiares, los datos obtenidos acerca de cómo el CEO había llegado a ese puesto fueron los siguientes:

Formó	Formó	Formó	Heredó
Heredó	Formó	Heredó	Formó
Heredó	Formó	Formó	Formó
Formó	Contrató	Contrató	Contrató
Heredó	Heredó	Heredó	Formó
Formó	Formó	Formó	Contrató
Formó	Heredó		

- a. Dé una distribución de frecuencias.
 b. Dé una distribución de frecuencias porcentuales.
 c. Presente una gráfica de barras.
 d. ¿Qué porcentaje de los CEOs de empresas familiares llegaron a ese puesto por heredar la empresa? ¿Cuál es la razón principal por la que una persona llega al puesto de CEO en una empresa familiar?
10. Netflix, Inc., de San José California, renta, por correo, más de 50 000 títulos de DVD. Los clientes ordenan en línea los DVDs que deseen ver. Antes de ordenar un DVD, el cliente puede ver una descripción del mismo y, si así lo desea, un resumen de las evaluaciones del mismo. Netflix emplea un sistema de evaluación de cinco estrellas que tienen el significado siguiente:

1 estrella	Me disgustó
2 estrellas	No me disgustó
3 estrellas	Me gustó
4 estrellas	Me gustó mucho
5 estrellas	Me fascinó

Dieciocho críticos, entre los que se encontraban Roger Ebert de *Chicago Sun Times* y Ty Burr de *Boston Globe*, proporcionaron evaluaciones en Hispanoamérica de la película *Batman inicia* (Netflix.com, 1 de marzo de 2006). Las evaluaciones fueron las siguientes:

- 4, 2, 5, 2, 4, 3, 3, 4, 4, 3, 4, 4, 4, 2, 4, 4, 5, 4
- Diga por qué son cualitativos estos datos.
 - Dé una distribución de frecuencias y una distribución de frecuencia relativa.
 - Dé una gráfica de barras.
 - Haga un comentario sobre las evaluaciones que dieron los críticos a esta película.

2.2

Resumen de datos cuantitativos

Distribución de frecuencia

TABLA 2.4

AUDITORÍA ANUAL
(DÍAS DE DURACIÓN)

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 de datos tabular que presenta el número de elementos (frecuencia) en cada una de las clases disyuntas. Esta definición es válida tanto para datos cualitativos como cuantitativos. Sin embargo, cuando se trata de datos cuantitativos se debe tener más cuidado al definir las clases disyuntas que se van a usar en la distribución de frecuencia.

Considere, por ejemplo, los datos cuantitativos de la tabla 2.4. En esta tabla se presenta la duración en días de una muestra de auditorías de fin de año de 20 clientes de una empresa pequeña de contadores públicos. Los tres pasos necesarios para definir las clases de una distribución de frecuencia con datos cuantitativos son

- Determinar el número de clases disyuntas.
- Determinar el ancho de cada clase
- Determinar los límites de clase.

Se mostrarán estos pasos elaborando una distribución de frecuencia con los datos de la tabla 2.4.

Número de clases Las clases se forman especificando los intervalos que se usarán para agrupar los datos. Se recomienda emplear entre 5 y 20 clases. Cuando los datos son pocos, cinco o seis clases bastan para resumirlos. Si son muchos, se suele requerir más clases. La idea es tener las clases suficientes para que se muestre la variación en los datos, pero no deben ser demasiadas si algunas de ellas contienen sólo unos cuantos datos. Como el número de datos en la tabla 2.4 es relativamente pequeño ($n = 20$), se decide elaborar una distribución de frecuencia con cinco clases.

Hacer las clases de una misma amplitud reduce la posibilidad de que los usuarios hagan interpretaciones inapropiadas.

Ancho de clase El segundo paso al construir una distribución de frecuencia para datos cuantitativos es elegir el ancho de las clases. Como regla general es recomendable que el ancho sea el mismo para todas las clases. Así, el ancho y el número de clases no son decisiones independientes. Entre mayor sea el número de clases menor es el ancho de las clases y viceversa. Para determinar el ancho de clase apropiada se empieza por identificar el mayor y el menor de los valores de los datos. Después, usando el número de clases deseado, se emplea la expresión siguiente para determinar el ancho aproximada de clase.

$$\text{Ancho aproximada de clase} = \frac{\text{Valor mayor en los datos} - \text{Valor menor en los datos}}{\text{Número de clase}} \quad (2.2)$$

El ancho aproximado de clase que se obtiene con la ecuación (2.2) se redondea a un valor más adecuado de acuerdo con las preferencias de la persona que elabora la distribución de frecuencia. Por ejemplo, si el ancho de clase aproximado es 9.28, se redondea a 10 porque 10 es un ancho de clase más adecuado para la presentación de la distribución de la frecuencia.

En los datos sobre las duraciones de las auditorías de fin de año el valor mayor en los datos es 33 y el valor menor es 12. Como se ha decidido resumir los datos en cinco clases, empleando



No hay una distribución de frecuencia que sea la mejor para un conjunto de datos. Distintas personas elaboran diferentes, pero igual de aceptables, distribuciones de frecuencia para un conjunto de datos dado. El objetivo es hacer notar el agrupamiento y la variación natural de los datos.

TABLA 2.5
DISTRIBUCIÓN DE FRECUENCIA DE LAS AUDITORÍAS

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

la ecuación (2.2) el ancho aproximado de clase que se obtiene es $(33 - 12)/5 = 4.2$. Por tanto, al redondear, en la distribución de frecuencia se usa como ancho de clase cinco días.

En la práctica el número de clases y su ancho adecuado se determinan por prueba y error. Una vez que se elige una determinado número de clases, se emplea la ecuación 2.2 para determinar el ancho aproximado de clase. El proceso se repite con distintos números de clases. El analista determina la combinación de número y ancho de clases que le proporciona la mejor distribución de frecuencia para resumir los datos.

En el caso de los datos de la tabla 2.4, una vez que se ha decidido emplear cinco clases, cada una con ancho de cinco días, el paso siguiente es especificar los límites de cada clase.

Límites de clase Los límites de clase deben elegirse de manera que cada dato pertenezca a una y sólo una de las clases. El *límite de clase inferior* indica el menor valor de los datos a que pertenece esa clase. El *límite de clase superior* indica el mayor valor de los datos a que pertenece esa clase. Al elaborar distribuciones de frecuencia para datos cualitativos, no es necesario especificar límites de clase porque cada dato corresponde de manera natural a una de las clases disyuntas. Pero con datos cuantitativos, como la duración de las auditorías de la tabla 2.4, los límites de clase son necesarios para determinar dónde colocar cada dato.

Mediante los datos de la duración de las auditorías de la tabla 2.4, se elige 10 días como límite inferior y 14 como límite superior de la primera clase. En la tabla 2.5, esta clase se denota como 10–14. El valor menor, 12 (de la tabla), pertenece a la clase 10–14. Después se elige 15 días como límite inferior y 19 como límite superior de la clase siguiente. Así, se continúan definiendo los límites inferior y superior de las clases hasta tener las cinco clases: 10–14, 15–19, 20–24, 25–29 y 30–34. El valor mayor en los datos, 33, pertenece a la clase 33–34. Las diferencias entre los límites inferiores de clase de clases adyacentes es el ancho de clase. Con los dos primeros límites inferiores de clase, 10 y 15, se ve que el ancho de clase es $15 - 10 = 5$.

Una vez determinados números, ancho y límites de las clases, la distribución de frecuencia se obtiene contando el número de datos que corresponden a cada clase. Por ejemplo, en la tabla 2.4 se observa que hay cuatro valores, 12, 14, 14 y 13, que 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 que se muestra en la tabla 2.5. En esta distribución de frecuencia se observa lo siguiente:

1. Las duraciones de las auditorías que se presentan con más frecuencia son de la clase 15–19 días. Ocho de las 20 auditorías caen en esta clase.
2. Sólo una auditoría requirió 30 o más días.

También se obtienen otras conclusiones, dependiendo de los intereses de quien observa la distribución de frecuencia. La utilidad de una distribución de frecuencia es que proporciona claridad acerca de los datos, la cual no es fácil de obtener con la forma desorganizada de éstos.

Punto medio de clase En algunas aplicaciones se desea conocer el punto medio de las clases de una distribución de frecuencia de datos cuantitativos. El **punto medio de clase** es el valor que queda a la mitad entre el límite inferior y el límite superior de la clase. En el caso de las duraciones de las auditorías, los cinco puntos medios de clase son 12, 17, 22, 27 y 32.

Distribuciones de frecuencia relativa y de frecuencia porcentual

Las distribuciones de frecuencia relativa y de frecuencia porcentual para datos cuantitativos se definen de la misma forma que para datos cualitativos. Primero debe recordar que la frecuencia relativa es el cociente, respecto al total de observaciones, de las observaciones que pertenecen a una clase. Si el número de observaciones es n ,

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

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 dado que $n = 20$, en la tabla 2.6 se muestran las distribuciones de frecuencia relativa y de frecuencia porcentual de los datos de las

TABLA 2.6 DISTRIBUCIONES DE FRECUENCIA RELATIVA Y DE FRECUENCIA PORCENTUAL CON LOS DATOS DE LAS DURACIONES DE LAS AUDITORÍAS

Duración de las auditorías (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

duraciones de las auditorías. Observe que 0.40 de las auditorías, o 40%, necesitaron entre 15 y 19 días. Sólo 0.05%, o 5%, requirió 30 o más días. De nuevo, hay más interpretaciones o ideas que se obtienen de la tabla 2.6.

Gráficas de puntos

Uno de los más sencillos resúmenes gráficos de datos son las **gráficas de puntos**. En el eje horizontal se presenta el intervalo de los datos. Cada dato se representa por un punto colocado sobre este eje. La figura 2.3 es la gráfica de puntos de los datos de la tabla 2.4. Los tres puntos que se encuentran sobre el 18 del eje horizontal indican que hubo tres auditorías de 18 días. Las gráficas de puntos muestran los detalles de los datos y son útiles para comparar la distribución de los datos de dos o más variables.

Histograma

Una presentación gráfica usual para datos cuantitativos es el **histograma**. Esta gráfica se hace con datos previamente resumidos mediante una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual. Un histograma se construye colocando la variable de interés en el eje horizontal y la frecuencia, la frecuencia relativa o la frecuencia porcentual en el eje vertical. La frecuencia, frecuencia relativa o frecuencia porcentual de cada clase se indica dibujando 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 las duraciones de las auditorías. Observe que la clase con mayor frecuencia se indica mediante el rectángulo que se encuentra sobre la clase 15–19 días. La altura del rectángulo muestra que la frecuencia de esta clase es 8. Un histograma de las distribuciones de frecuencia relativa o porcentual de estos datos se ve exactamente igual que el histograma de la figura 2.4, excepto que en el eje vertical se colocan los valores de frecuencia relativa o porcentual.

Como se muestra en la figura 2.4, los rectángulos adyacentes de un histograma se tocan uno a otro. A diferencia de las gráficas de barras, en un histograma no hay una separación natural en

FIGURA 2.3 GRÁFICA DE PUNTOS PARA LOS DATOS DE LAS DURACIONES DE LAS AUDITORÍAS

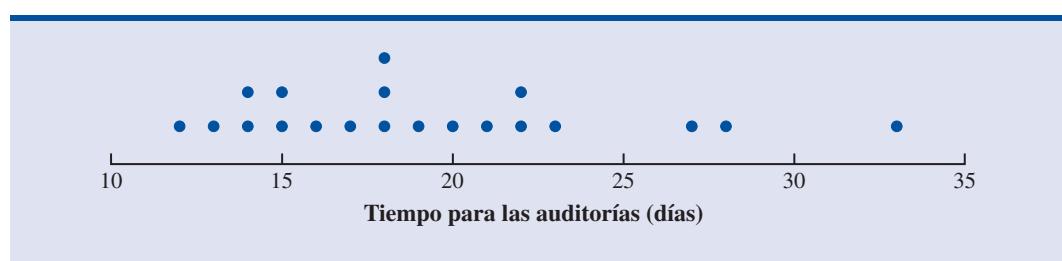
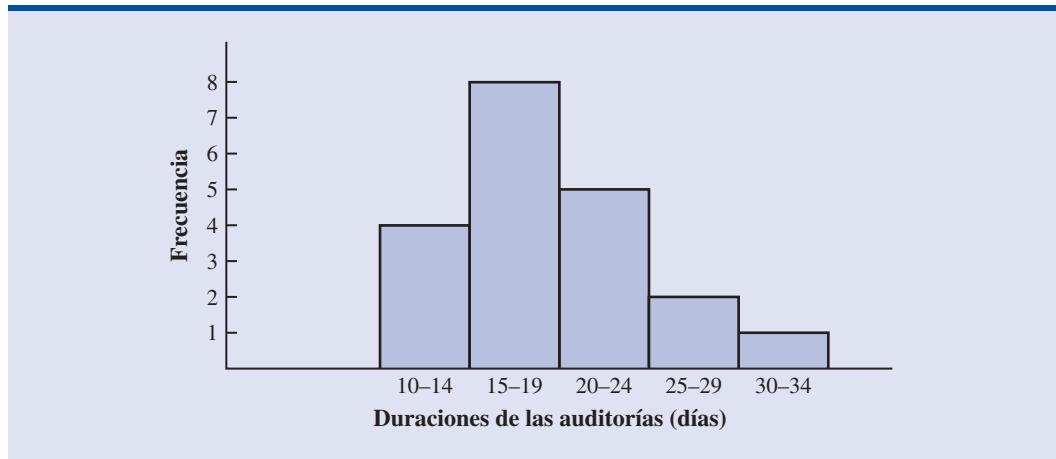


FIGURA 2.4 HISTOGRAMA DE LOS DATOS DE LAS DURACIONES DE LAS AUDITORÍAS

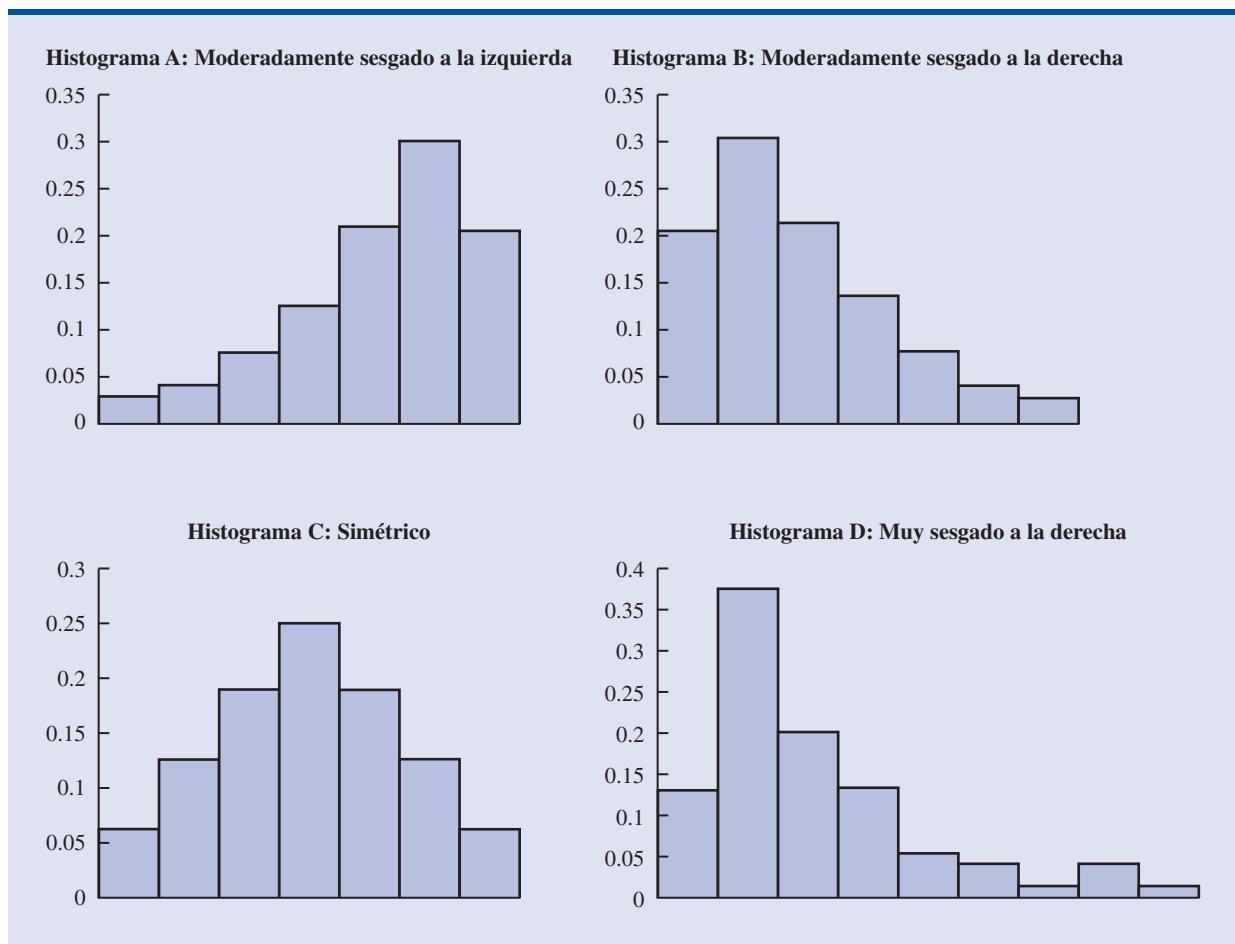
tre los rectángulos de clases adyacentes. Este formato es el usual para histogramas. Como las clases de las duraciones de las auditorías son 10–14, 15–19, 20–24, 25–29 y 30–34 parecería que se necesitará una unidad de espacio entre las clases, de 14 a 15, de 19 a 20, de 24 a 25 y de 29 a 30. Cuando se construye un histograma se eliminan estos espacios. Eliminar los espacios entre las clases del histograma de las duraciones de las auditorías sirve para indicar que todos los valores entre el límite inferior de la primera clase y el superior de la última son posibles.

Uno de los usos más importantes de un histograma es proveer información acerca de la forma de la distribución. En la figura 2.5 se muestran cuatro histogramas construidos a partir de distribuciones de frecuencia relativa. En el histograma A se muestra un conjunto de datos moderadamente sesgado a la izquierda. Se dice que un histograma es sesgado a la izquierda si su cola se extiende más hacia la izquierda. Dichos histogramas son típicos para calificaciones: no hay calificaciones mayores a 100%, la mayor parte están arriba de 70% y sólo hay unas cuantas bajas. En el histograma B se muestra un conjunto de datos moderadamente sesgado a la derecha. Un histograma está sesgado a la derecha si su cola se extiende más hacia la derecha. Ejemplos de este tipo de histogramas son los datos de los precios de las casas; unas cuantas casas caras crean el sesgo a la derecha.

En C se observa un histograma simétrico. En éste la cola izquierda es la imagen de la cola derecha. Los histogramas de datos para aplicaciones nunca son perfectamente simétricos, pero en muchas aplicaciones suelen ser más o menos simétricos. En D se observa un histograma muy sesgado a la derecha. Éste se elaboró con datos sobre la cantidad de compras a lo largo de un día en una tienda de ropa para mujeres. Los datos de aplicaciones de negocios o economía suelen conducir a histogramas sesgados a la derecha. Por ejemplo datos de los precios de las casas, de los salarios, de las cantidades de las compras, etc., suelen dar histogramas sesgados a la derecha.

Distribuciones acumuladas

Una variación de las distribuciones de frecuencia que proporcionan otro resumen tabular de datos cuantitativos es la **distribución de frecuencia acumulada**. La distribución de frecuencia acumulada usa la cantidad, las amplitudes y los límites de las clases de la distribución de frecuencia. Sin embargo, en lugar de mostrar la frecuencia de cada clase, la distribución de frecuencia acumulada muestra la cantidad de datos que tienen un valor *menor o igual* al límite superior de cada clase. Las primeras dos columnas de la tabla 2.7 corresponden a la distribución de frecuencia acumulada de los datos de las duraciones de las auditorías.

FIGURA 2.5 HISTOGRAMAS CON DISTINTOS TIPOS DE SESGO

Para entender cómo se determina la frecuencia acumulada, considere la clase que dice “menor o igual que 24”. La frecuencia acumulada en esta clase es simplemente la suma de la frecuencia de todas las clases en que los valores de los datos son menores o iguales que 24. En la distribución de frecuencia de la tabla 2.5 la suma de las frecuencias para las clases 10–14, 15–29 y 20–24 indica que los datos cuyos valores son menores o iguales que 24 son $4 + 8 + 5 = 17$. Por lo tanto, en esta clase la frecuencia acumulada es 17. Además, en la distribución de frecuen-

TABLA 2.7 DISTRIBUCIONES DE FRECUENCIA ACUMULADA, FRECUENCIA RELATIVA ACUMULADA Y FRECUENCIA PORCENTUAL ACUMULADA

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

cias acumuladas de la tabla 2.7 se observa que cuatro auditorías duraron 14 días o menos y que 19 auditorías duraron 29 días o menos.

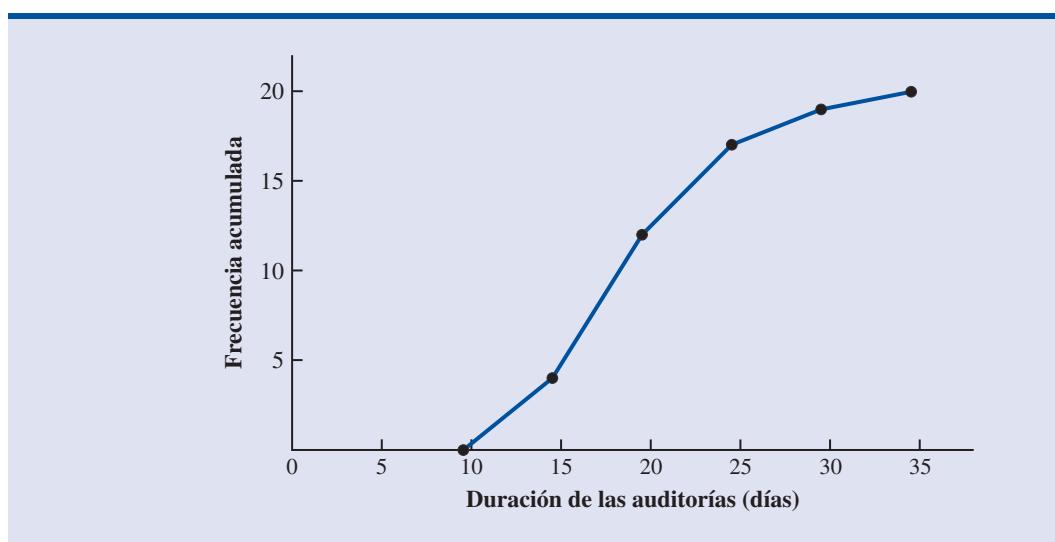
Por último, se tiene que la **distribución de frecuencias relativas acumuladas** indica la proporción de todos los datos que tienen valores menores o iguales al límite superior de cada clase, y la **distribución de frecuencias porcentuales acumuladas** indica el porcentaje de todos los datos que tienen valores menores o iguales al límite superior de cada clase. La distribución de frecuencias relativas acumuladas se calcula ya sea sumando las frecuencias relativas que aparecen en la distribución de frecuencias relativas o dividiendo la frecuencia acumulada entre la cantidad total de datos. Empleando el último método, las frecuencias relativas acumuladas que aparecen en la columna 3 de la tabla 2.7 se obtienen dividiendo las frecuencias acumuladas de la columna 2 entre la cantidad total de datos ($n = 20$). Las frecuencias porcentuales acumuladas se obtienen multiplicando las frecuencias relativas por 100. Estas distribuciones de frecuencias acumuladas relativas y porcentuales indican que 0.85 o el 85% de las auditorías se realizaron en 24 días o menos, 0.95 o 95% de las auditorías se realizaron en 29 días o menos, etcétera.

Ojiva

La gráfica de una distribución acumulada, llamada **ojiva**, es una gráfica que muestra los valores de los datos en el eje horizontal y las frecuencias acumuladas, las frecuencias relativas acumuladas o las frecuencias porcentuales acumuladas en el eje vertical. En la figura 2.6 se muestra una ojiva correspondiente a las frecuencias acumuladas de las duraciones de las auditorías.

La ojiva se construye al graficar cada uno de los puntos correspondientes a la frecuencia acumulada de las clases. Como las clases de las duraciones de las auditorías son 10–14, 15–19, 20–24, etc., hay huecos de una unidad entre 14 y 15, 19 y 20, etc. Estos huecos se eliminan al graficar puntos a la mitad entre los dos límites de clase. Así, para la clase 10–14 se usa 14.5, para la clase 15–19 se usa 19.5 y así en lo sucesivo. En la ojiva de la figura 2.6 la clase “menor o igual que 14” cuya frecuencia acumulada es 4 se grafica mediante el punto que se localiza a 14.5 unidades sobre el eje horizontal y a 4 unidades sobre el vertical. La clase “menor o igual que 19” cuya frecuencia acumulada es 12 se representa por un punto que se encuentra a 19.5 unidades sobre el eje horizontal y 12 unidades sobre el vertical. Observe que en el extremo izquierdo de la ojiva se ha graficado un punto más. Este punto inicia la ojiva mostrando que en los datos no hay valores que se encuentren abajo de la clase 10–14. Este punto se encuentra a 9.5 unidades sobre el eje horizontal y a 0 unidades sobre el vertical. Para terminar los puntos graficados se conectan mediante líneas rectas.

FIGURA 2.6 OJIVA DE LOS DATOS DE LAS DURACIONES DE LAS AUDITORÍAS



NOTAS Y COMENTARIOS

1. Una gráfica de barras y un histograma son en esencia lo mismo; ambas son representaciones gráficas de una distribución de frecuencia. Un histograma es sólo una gráfica de barras sin separación entre las barras. Para algunos datos cuantitativos discretos, también se puede tener separación entre las barras. Considere por ejemplo, el número de materias en que está inscrito un estudiante universitario. Los datos sólo tienen valores enteros. No hay valores intermedios como 1.5, 2.73, etc. Sin embargo cuando se tienen datos cuantitativos continuos, como en las auditorías, no es apropiado tener separación entre las barras.
2. Los valores adecuados para los límites de clase cuando se tienen datos cuantitativos depende del nivel de precisión de los datos. Por ejemplo, en el caso de los datos de la tabla 2.4, sobre la duración de las auditorías, los límites usados fueron números enteros. Si los datos hubieran estado redondeados a la décima de día más cercana (es decir, 12.3, 14.4, etc.), entonces los límites se hubieran dado con décimas de día. La primera clase, por ejemplo, hubiera sido de 10.0 a 14.9. Si los datos se hubieran registrado hasta la centésima de día más cercana (es decir, 12.34, 14.45, etc.), los límites se hubieran dado con centésimas de días. Por ejemplo la primera clase hubiera sido de 10.00–14.99.
3. Una clase *abierta* sólo necesita el límite inferior de la clase o el límite superior de la clase. Por ejemplo, suponga que en los datos de la tabla 2.4 sobre las duraciones de las auditorías dos de éstas hubieran durado 58 y 65 días. En lugar de haber seguido con clases de amplitud 5 de 35–39, de 40–44, de 45 a 49, etc., podría haber simplificado la distribución de frecuencia mediante una clase abierta de “35 o más”. La frecuencia de esta clase habría sido 2. La mayor parte de las clases abiertas aparecen en el extremo superior de la distribución. Algunas veces se encuentran clases abiertas en el extremo inferior y rara vez están en ambos extremos.
4. En una distribución de frecuencia acumulada, la última frecuencia siempre es igual al número total de observaciones. En una distribución de frecuencia relativa acumulada la última frecuencia siempre es igual a 1.00 y en una distribución de frecuencia porcentual acumulada la última frecuencia es siempre 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.
 - Elabore una distribución de frecuencia relativa y una de frecuencia porcentual usando las clases del inciso a.
12. Considere la distribución de frecuencia siguiente.

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

Construya una distribución de frecuencia acumulada y otra de frecuencia relativa acumulada.



13. Con los datos del ejercicio 12 elabore un histograma y una ojiva.
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

- Construya un diagrama de punto.
- Elabore una distribución de frecuencia.
- Construya una distribución de frecuencia porcentual.

Aplicaciones



15. El personal de un consultorio analiza los tiempos de espera de los pacientes que requieren servicio de emergencia. Los datos siguientes son los tiempos de espera en minutos recolectados a lo largo de un mes.

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

Con las clases 0–4, 5–9, etcétera.

- Muestre la distribución de la frecuencia.
 - Exprese la distribución de la frecuencia relativa.
 - Muestre la distribución de frecuencia acumulada.
 - Presente la distribución de frecuencia relativa acumulada.
 - ¿Cuál es la proporción de los pacientes que requieren servicio de emergencia y esperan 9 minutos o menos?
16. Considere las dos distribuciones de frecuencias siguientes. La primera distribución de frecuencia proporciona el ingreso anual bruto ajustado de Estados Unidos (Internal Revenue Service, marzo 2003). La segunda distribución de frecuencia muestra las calificaciones de exámenes de un grupo de estudiantes universitarios en un curso de estadística.

Ingreso (en miles de \$)	Frecuencia (en millones)	Calificaciones de examen	Frecuencia
0–24	60	20–29	2
25–49	33	30–39	5
50–74	20	40–49	6
75–99	6	50–59	13
100–124	4	60–69	32
125–149	2	70–79	78
150–174	1	80–89	43
175–199	1	90–99	21
Total	127	Total	200

- Con los datos del ingreso anual elabore un histograma. ¿Qué evidencia de sesgo observa? ¿Es razonable este sesgo? Explique.
 - Con los datos de las calificaciones elabore un histograma. ¿Qué evidencia de sesgo observa? Explique.
 - Con los datos del ejercicio 11 elabore un histograma. ¿Qué evidencia de sesgo observa? ¿Cuál es la forma general de la distribución?
17. ¿Cuál es el precio típico de las acciones de las 30 empresas del promedio industrial Dow Jones? Los datos siguientes son los precios de las acciones, al dólar más cercano, en enero de 2006 (*The Wall Street Journal*, 16 de enero de 2006).

archivo
en
PriceShare

Empresa	\$/Acción	Empresa	\$/Acción
AIG	70	Home Depot	42
Alcoa	29	Honeywell	37
Altria Group	76	IBM	83
American Express	53	Intel	26
AT&T	25	Johnson & Johnson	62
Boeing	69	JPMorgan Chase	40
Caterpillar	62	McDonald's	35
Citigroup	49	Merck	33
Coca-Cola	41	Microsoft	27
Disney	26	3M	78
DuPont	40	Pfizer	25
ExxonMobil	61	Procter & Gamble	59
General Electric	35	United Technologies	56
General Motors	20	Verizon	32
Hewlett-Packard	32	Wal-Mart	45

- a. Con estos datos elabore una distribución de frecuencia.
- b. Con estos datos elabore un histograma. Interprete el histograma, presente un análisis de la forma general del histograma, el precio medio de cada intervalo de acciones, el precio más frecuente por intervalo de acciones, los precios más alto y más bajo por acción.
- c. ¿Cuáles son las acciones que tienen el precio más alto y el más bajo?
- d. Use *The Wall Street Journal* para encontrar los precios actuales por acción de estas empresas. Elabore un histograma con estos datos y discuta los cambios en comparación con enero de 2006.
18. NRF/BIG proporciona los resultados de una investigación sobre las cantidades que gastan en vacaciones los consumidores (*USA Today*, 20 de diciembre de 2005). Los datos siguientes son las cantidades gastadas en vacaciones por los 25 consumidores de una muestra.

archivo
en
Holiday

1200	850	740	590	340
450	890	260	610	350
1780	180	850	2050	770
800	1090	510	520	220
1450	280	1120	200	350

- a. ¿Cuál es la menor cantidad gastada en vacaciones? ¿Cuál la mayor?
- b. Use \$250 como amplitud de clase para elaborar con estos datos una distribución de frecuencia y una distribución de frecuencia porcentual.
- c. Elabore un histograma y comente la forma de la distribución.
- d. ¿Qué observaciones le permiten hacer las cantidades gastadas en vacaciones?
19. El correo no deseado afecta la productividad de los oficinistas. Se hizo una investigación con oficinistas para determinar la cantidad de tiempo por día que pierden en estos correos no deseados. Los datos siguientes corresponden a los tiempos en minutos perdidos por día observados en una muestra.

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

Resuma estos datos construyendo:

- a. Una distribución de frecuencia (con las clases 1–5, 6–10, 11–15, 16–20, etc.)
- 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 los oficinistas pierde 5 minutos o menos en revisar el correo no deseado? ¿Qué porcentaje pierde más de 10 minutos por día en esto?
20. A continuación se presentan las 20 mejores giras de concierto y el precio promedio del costo de sus entradas en Estados Unidos. Esta lista se basa en datos proporcionados por los promotores y administradores de los locales a la publicación *Pollstar* (*Associated Press*, 21 de noviembre de 2003).



Concerts

Gira de conciertos	Precio de la entrada	Gira de conciertos	Precio de la entrada
Bruce Springsteen	\$72.40	Toby Keith	\$37.76
Dave Matthews Band	44.11	James Taylor	44.93
Aerosmith/KISS	69.52	Alabama	40.83
Shania Twain	61.80	Harper/Johnson	33.70
Fleetwood Mac	78.34	50 Cent	38.89
Radiohead	39.50	Steely Dan	36.38
Cher	64.47	Red Hot Chili Peppers	56.82
Counting Crows	36.48	R.E.M.	46.16
Timberlake/Aguilera	74.43	American Idols Live	39.11
Mana	46.48	Mariah Carey	56.08

- Resuma los datos construyendo:
- a. Una distribución de frecuencia y una distribución de frecuencia porcentual.
 - b. Un histograma.
 - c. ¿Qué concierto tiene el precio promedio más alto? ¿Qué concierto tiene el precio promedio menos caro?
 - d. Haga un comentario sobre qué indican los datos acerca de los precios promedio de las mejores giras de concierto.
21. *Nielsen Home Technology Report* informa sobre la tecnología en el hogar y su uso. Los datos siguientes son las horas de uso de computadora por semana en una muestra de 50 personas.



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 estos datos construyendo:
- a. Una distribución de frecuencia (como ancho de clase use tres horas).
 - b. Una distribución de frecuencia relativa.
 - c. Un histograma.
 - d. Una ojiva.
 - e. Haga un comentario sobre lo que indican los datos respecto al uso de la computadora en el hogar.

2.3

Análisis exploratorio de datos: el diagrama de tallo y hojas

Las técnicas del **análisis exploratorio de datos** emplean aritmética sencilla y gráficas fáciles de dibujar útiles para resumir datos. La técnica conocida como **diagrama de tallo y hojas** muestra en forma simultánea el orden jerárquico y la forma de un conjunto de datos.

TABLA 2.8 NÚMERO DE PREGUNTAS CONTESTADAS CORRECTAMENTE EN UN EXAMEN 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

Para ilustrar el uso de los diagramas de tallo y hojas, considere la tabla 2.8. Estos datos son el resultado de un examen de aptitudes con 150 preguntas presentado por 50 personas que aspiraban a un puesto en una empresa. Los datos indican el número de respuestas correctas por examen.

Para elaborar un diagrama de tallo y hoja inicie acomodando los primeros dígitos de cada uno de los datos a la izquierda de una línea vertical. A la derecha de la línea vertical se anota el último dígito de cada dato. Con base en el primer renglón de la tabla 2.8 (112, 72, 69, 97 y 107), los primeros cinco datos al elaborar el diagrama de tallo y hojas serían los siguientes:

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

Por ejemplo, para el dato 112, se observa que los primeros dígitos, 11, se encuentran a la izquierda de la línea y el último dato, 2, a la derecha. De manera similar, el primer dígito, 7, del dato 72 se encuentra a la izquierda de la línea y el 2 a la derecha. Si continúa colocando el último dígito de cada dato en el renglón correspondiente a sus primeros dígitos obtiene:

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										

Una vez organizados los datos de esta manera, ordenar los datos de cada renglón de menor a mayor es sencillo. Entonces obtiene el diagrama de tallo y hojas que se muestra 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 vertical es una *hoja*. Por ejemplo, considere el primer renglón que tiene como tallo el 6 y como hojas 8 y 9.

6 | 8 9

Este renglón indica que hay dos datos que tienen como primer dígito el seis. Las hojas indican que estos datos son 68 y 69. De manera similar, el segundo renglón

7 | 2 3 3 5 6 6

indica que hay seis datos que tienen como primer dígito el 7. Las hojas indican que estos datos son 72, 73, 73, 75, 76 y 76.

Para atender a la forma del diagrama de tallo y hojas, se usan rectángulos que contienen las hojas de cada tallo; con esto 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										

Al rotar la página sobre su costado en contra de las manecillas del reloj se obtiene una imagen de los datos que es parecida a un histograma y en el que las clases son 60–69, 70–79, 80–89, etcétera.

Aunque el diagrama de tallo y hojas parece proporcionar la misma información que un histograma, tiene dos ventajas fundamentales.

1. El diagrama de tallo y hojas es más fácil de construir a mano.
2. En cada intervalo de clase proporciona más información que un histograma debido a que el tallo y la hoja proporcionan el dato.

Así como para una distribución de frecuencia o para un histograma no hay un determinado número de clases, tampoco para el diagrama de tallo y hojas hay un número determinado de renglones a tallos. Si piensa que el diagrama de tallo y hojas original condensa demasiado los datos, es fácil expandirlo empleando dos o más tallos por cada primer dígito. Por ejemplo, para usar

En un diagrama expandido de tallo y hojas, siempre que un tallo aparece dos veces, al primero le corresponden las hojas 0–4 y al segundo las hojas 5–9.

dos tallos por cada primer dígito se ponen todos los datos que terminen en 0, 1, 2, 3 o 4 en un renglón y todos los datos que terminen en 5, 6, 7, 8 o 9 en otro. Este método se ilustra en el siguiente diagrama expandido de tallo y hojas.

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 las hojas de los datos 72, 73 y 73 pertenecen al intervalo 0–4 y aparecen con el primer tallo que tiene el valor 7. Las hojas de los valores 75, 76 y 76 pertenecen al intervalo 5–9 y aparecen con el segundo tallo que tiene el valor 7. Este diagrama expandido de tallo y hojas es semejante a una distribución con los intervalos 65–69, 70–74, 75–79, etcétera.

El ejemplo anterior muestra un diagrama de tallo y hojas con datos de hasta tres dígitos. Estos diagramas también se elaboran con datos de más de tres dígitos. Por ejemplo, considere los datos siguientes sobre el número de hamburguesas vendidas en un restaurante de comida rápida en cada una de 15 semanas.

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

A continuación se presenta un diagrama de tallo y hojas de estos datos.

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

En un diagrama de tallo y hojas se usa un solo dígito para definir cada hoja. La unidad de hoja indica por qué número debe multiplicar los números del tallo y la hoja para aproximar el dato original. Las unidades de hoja son 100, 10, 1, 0.1 etcétera.

Observe que para definir cada hoja se emplea un solo dígito y que para construir el diagrama sólo se usaron los primeros tres dígitos de cada dato. En la parte superior del diagrama se ha especificado que la Unidad de hoja = 10. Para ilustrar cómo se interpretan los datos de este diagrama considere el primer tallo 15 y su hoja correspondiente 6. Al unir estos números obtiene 156. Para lograr una aproximación al dato original es necesario multiplicar este número por 10, el valor de la *unidad de hoja*. Por tanto, $156 \times 10 = 1560$ es una aproximación al dato original empleado para construir el diagrama de tallo y hoja. Aunque a partir de este diagrama no es posible reconstruir los datos exactos, la convención de usar un solo dígito para cada hoja, permite construir diagramas de tallo y hojas con datos que tengan un gran número de dígitos. En diagramas de tallo y hojas en los que no se especifica la unidad de hoja, se supone que la unidad es 1.

Ejercicios

Métodos

22. Con los datos siguientes construya un diagrama de tallo y hojas.

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

23. Con los datos siguientes construya un diagrama de tallo y hojas.

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. Con los datos siguientes construya un diagrama de tallo y hojas. Use 10 como unidad de hoja.

1161	1206	1478	1300	1604	1725	1361	1422
1221	1378	1623	1426	1557	1730	1706	1689



Aplicaciones



25. Un psicólogo elabora una nueva prueba de inteligencia para adultos. Aplica la prueba a 20 individuos y obtiene los datos siguientes.

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

Construya un diagrama de tallo y hojas.

26. La asociación estadounidense de inversionistas individuales realiza una investigación anual sobre intermediarios de descuento. Las siguientes son las comisiones en una muestra de 24 intermediarios (*AAII Journal*, enero de 2003). Estas son dos tipos de operaciones con asistencia de 100 acciones a \$50 cada una y una operación en línea de 500 acciones a \$50 cada una.



Corredor	Operación con asistencia de 100 acciones \$50/ acción	Operación en línea de 500 acciones a \$50 /acción	Corredor	Operación con asistencia de 100 acciones \$50/ acción	Operación en línea de 500 acciones a \$50/acción
Accutrade	30.00	29.95	Merrill Lynch Direct	50.00	29.95
Ameritrade	24.99	10.99	Muriel Siebert	45.00	14.95
Banc 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

- Redondee los precios al dólar más cercano y elabore un diagrama de tallo y hojas de las 100 acciones a \$50 por acción. Haga un comentario sobre la información que obtuvo acerca de estos precios.
 - Redondee los precios al dólar más cercano y elabore un diagrama de tallo y hojas de las 500 acciones a \$50 por acción. Haga un comentario sobre estos precios.
27. La mayor parte de los centros turísticos importantes de esquí de Estados Unidos ofrecen programas familiares con clases de esquí para niños. Por lo general proporcionan 4 a 6 horas de clase con un instructor certificado. A continuación se presentan las cuotas diarias en 15 centros turísticos. (*The Wall Street Journal*, 20 de enero de 2006).

Centro turístico	Ubicación	Cuota diaria	Centro turístico	Ubicación	Cuota 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	New 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. Con estos datos elabore un diagrama de tallo y hojas.
- b. Interprete el diagrama de tallo y hojas en términos de lo que expresa de las cuotas diarias de estos programas.
28. Para un maratón (13.1 millas) en Florida en 2004 hubo 1228 registrados (*Naples Daily News*, 17 de enero de 2004). Para esta competencia hubo seis grupos de edades. Los datos siguientes son las edades encontradas en una muestra de 40 participantes.

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. Realice un diagrama expandido de tallo y hojas.
- b. ¿En qué grupo de edad hubo más participantes?
- c. ¿Qué edad se presenta con más frecuencia?
- d. En un artículo del *Naples Daily News* se hace énfasis sobre la cantidad de corredores de veintitantes años. ¿Qué porcentaje de los corredores pertenecían al grupo de veintitantes años? ¿Cuál supone qué era el tema del artículo?



2.4 Tabulaciones cruzadas y diagramas de dispersión

Las tabulaciones cruzadas y los diagramas de dispersión son empleados para presentar un resumen de datos, de tal manera que revele la relación entre las dos variables.

Este capítulo, hasta ahora, se ha concentrado en los métodos tabulares y gráficos empleados para resumir datos de una *sola variable*. Con frecuencia, los directivos o quienes deben tomar decisiones requieren métodos tabulares o gráficos que les ayuden a entender 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 tabular de los datos de dos variables. El uso de la tabulación cruzada se ilustrará con los datos de la aplicación siguiente, que se basan en datos de *Zagat's Restaurant Review*. Se recolectaron los datos correspondientes a la calidad y precios de 300 restaurantes en el área de Los Ángeles. La tabla 2.9 muestra los datos de los 10 primeros restaurantes. Se presentan los datos de calidad y precio característicos de estos restaurantes. La calidad es una variable cualitativa que tiene como categorías bueno, muy bueno y excelente. El precio es una variable cuantitativa que va desde \$10 hasta \$49.

En la tabla 2.10 se muestra una tabulación cruzada con los datos de esta aplicación. El encabezado de la primera columna y el primer renglón definen las clases para las dos variables. Los encabezados de los renglones en el margen izquierdo (buena, muy buena y excelente) corresponden a las tres categorías de calidad. Los encabezados de las columnas (\$10–19, \$20–29, \$30–39 y

TABLA 2.9 EVALUACIÓN DE LA CALIDAD Y PRECIOS DE 300 RESTAURANTES DE LOS ÁNGELES

Restaurante	Calidad	Precio
1	Bueno	18
2	Muy bueno	22
3	Bueno	28
4	Excelente	38
5	Muy bueno	33
6	Bueno	28
7	Muy bueno	19
8	Muy bueno	11
9	Muy bueno	23
10	Bueno	13
.	.	.
.	.	.
.	.	.

\$40–49) corresponden a las cuatro clases de la variable precio. Para cada restaurante de la muestra se tiene el nivel de calidad y el precio. Por tanto, a cada restaurante de la muestra le corresponde una celda en un renglón y en una columna de la tabla. Por ejemplo, si el restaurante 5 tiene muy buena calidad y su precio es \$33, a este restaurante le corresponde el renglón 2 y la columna 3 de la tabla 2.10. Así que para elaborar una tabulación cruzada, simplemente se cuenta el número de restaurantes que pertenecen a cada una de las celdas de la tabla de tabulación cruzada.

La tabla 2.10 muestra que la mayor parte de los restaurantes de la muestra (64) tienen muy buena calidad y su precio está en el intervalo \$20–29. También se ve que sólo dos restaurantes tienen una calidad excelente y un precio en el intervalo \$10–19. Así es posible hacer interpretaciones semejantes con el resto de las frecuencias. Observe además que en el margen derecho y en el renglón inferior de la tabulación cruzada aparecen las distribuciones de frecuencia de la calidad y de los precios, por separado. En la distribución de frecuencia de la calidad, en el margen derecho, se observa que hay 84 restaurantes buenos, 150 muy buenos y 66 restaurantes excelentes. De manera semejante, en el renglón inferior se tiene la distribución de frecuencia de la variable precios.

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

Calidad	Frecuencia relativa	Frecuencia porcentual
Bueno	0.28	28
Muy bueno	0.50	50
Excelente	0.22	22
Total	1.00	100

TABLA 2.10 TABULACIÓN CRUZADA DE CALIDAD Y PRECIO DE 300 RESTAURANTES DE LOS ÁNGELES

Calidad	Precio				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
Total	78	118	76	28	300

En esta distribución de frecuencia porcentual se observa que 28% de los restaurantes son calificados como buenos, 50% como muy buenos y 22% excelentes.

Si divide los totales del renglón inferior de la tabulación cruzada entre el total de ese renglón obtiene distribuciones de frecuencia relativa y de frecuencia porcentual de los precios.

Precio	Frecuencia relativa	Frecuencia porcentual
\$10–19	0.26	26
\$20–29	0.39	39
\$30–39	0.25	25
\$40–49	0.09	9
Total	1.00	100

Observe que la suma de los valores en cada columna no tiene correspondencia exacta con el total de la columna debido a que los valores que se suman han sido redondeados. En esta distribución de frecuencia porcentual 26% de los precios se encuentran en la clase de los precios más bajos, 39% se encuentran en la clase siguiente, etcétera.

Las distribuciones de frecuencia y de frecuencia relativa obtenidas de los márgenes de las tabulaciones cruzadas proporcionan información de cada una de las variables por separado, pero no dan ninguna luz acerca de la relación entre las variables. El principal valor de una tabulación cruzada es que permite ver la relación entre las variables. Una observación de la tabulación cruzada de la tabla 2.10 es que los precios más altos están relacionados con la mejor calidad de los restaurantes y los precios bajos están relacionados con menor calidad.

Si se convierten las cantidades de una tabulación cruzada en porcentajes de columna o de renglón, se obtiene más claridad sobre la relación entre las variables. En la tabla 2.11 se presentan los porcentajes de renglón, que son el resultado de dividir cada frecuencia de la tabla 2.10 entre el total del renglón correspondiente. Entonces, cada renglón de la tabla 2.11 es una distribución de frecuencia porcentual de los precios en esa categoría de calidad. Entre los restaurantes de menor calidad (buenos), el mayor porcentaje corresponde a los menos caros (50% tiene precios en el intervalo \$10–19 y 47.6% en el intervalo \$20–29). De los restaurantes de mayor calidad (excelentes), los porcentajes mayores corresponden a los más caros (42.4% tiene precios de \$30–39 y 33.4% de \$40–49). Así que un precio más elevado está relacionado con una mejor calidad de los restaurantes.

La tabulación cruzada se utiliza mucho para examinar la relación entre dos variables. En la práctica, los informes finales de muchos estudios estadísticos contienen una gran cantidad de tabulaciones cruzadas. En este estudio sobre los restaurantes de Los Ángeles, en la tabulación cruzada se emplea una variable cualitativa (las calidades) y una cuantitativa (los precios). También se elaboran tabulaciones cruzadas con dos variables cualitativas o cuantitativas. Cuando se usan variables cuantitativas, primero es necesario crear las clases para los valores de las variables. Por ejemplo, en el caso de los restaurantes se agruparon los precios en cuatro categorías (\$10–19, \$20–29, \$30–39 y \$40–49).

TABLA 2.11 PORCENTAJES DE RENGLÓN DE CADA CATEGORÍA DE CALIDAD

Calidad	Precio				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

Paradoja de Simpson

Es posible combinar o agregar los datos de dos o más tabulaciones cruzadas para obtener una tabulación cruzada resumida que muestre la relación entre dos variables. En tales casos hay que tener mucho cuidado al sacar conclusiones acerca de la relación entre las dos variables de la tabulación cruzada agregada. En algunos casos las conclusiones obtenidas de la tabulación cruzada agregada se invierten por completo al observar los datos no agregados, situación conocida como **paradoja de Simpson**. Para ilustrar la paradoja de Simpson, se proporciona un ejemplo en el que se analizan las sentencias de dos jueces en dos tipos de tribunales.

Los jueces Ron Luckett y Dennis Kendall, presidieron los tres últimos años dos tipos de tribunales, de primera instancia y municipal. Algunas de las sentencias por ellos dictadas fueron apeladas. En la mayor parte de los casos los tribunales de apelación ratificaron las sentencias, pero en algunos casos fueron revocadas. Para cada juez se elabora una tabulación cruzada con las variables: sentencia (ratificada o revocada) y tipo de tribunal (de primera instancia y municipal). Suponga que después se combinan las dos tabulaciones cruzadas agregando los datos de los dos tipos de tribunales. La tabulación cruzada agregada que se obtiene tiene dos variables: sentencia (ratificada o revocada) y juez (Luckett o Kendall). En esta tabulación cruzada para cada uno de los jueces se da la cantidad de sentencias que fueron ratificadas y la cantidad de sentencias que fueron revocadas. En la tabla siguiente se presentan estos resultados junto a los porcentajes de columna entre paréntesis al lado de cada valor.

		Juez		
		Luckett	Kendall	Total
Sentencia	Ratificada	129 (86%)	110 (88%)	239
	Revocada	21 (14%)	15 (12%)	36
	Total (%)	150 (100%)	125 (100%)	275

Al analizar la columna de porcentajes resulta que 14% de las sentencias del juez Luckett fueron revocadas, pero del juez Kendall sólo 12% de las sentencias lo fueron. Por tanto, el juez Kendall tuvo un mejor desempeño, ya que de sus sentencias se ratificó un porcentaje mayor. Sin embargo, de esta conclusión surge un problema.

En la tabla siguiente se muestran los casos atendidos por cada uno de los jueces en los dos tribunales; aquí también se dan los porcentajes entre paréntesis al lado de los valores.

Juez Luckett				Juez Kendall			
Sentencia	Tribunal de primera instancia	Tribunal municipal	Total	Sentencia	Tribunal de primera instancia	Tribunal municipal	Total
Ratificada	29 (91%)	100 (85%)	129	Ratificada	90 (90%)	20 (80%)	110
Revocada	3 (9%)	18 (15%)	21	Revocada	10 (10%)	5 (20%)	15
Total (%)	32 (100%)	118 (100%)	150	Total (%)	100 (100%)	25 (100%)	125

Respecto de los porcentajes de Luckett, en el tribunal de primera instancia 91% de sus sentencias fueron ratificadas y en el tribunal municipal 85% lo fueron. En cuanto a los porcentajes de Kendall, 90% de sus sentencias del tribunal de primera instancia y 80% del tribunal municipal fueron ratificadas. Al comparar los porcentajes de columna de los dos jueces, es obvio que el juez Luckett tuvo un mejor desempeño en ambos tribunales que el Juez Kendall. Esto contradice las conclusiones obtenidas al agregar los datos de los dos tribunales en la primera tabulación cruzada. Se pensó que el juez Kendall tenía un mejor desempeño. Este ejemplo ilustra la paradoja de Simpson.

La primera tabulación cruzada se obtuvo agregando los datos de los dos tribunales de dos tabulaciones cruzadas. Observe que los dos jueces tuvieron porcentajes mayores de sentencias revocadas en las sentencias del tribunal municipal que en las del tribunal de primera instancia. Como el juez Luckett tuvo un porcentaje mayor de casos del tribunal municipal, los datos agregados favorecieron al juez Kendall. Sin embargo, si presta atención a las tabulaciones cruzadas de cada uno de los jueces, es claro que el juez Luckett tuvo un mejor desempeño. Por tanto, en la primera tabulación cruzada el *tipo de tribunal* es una variable oculta que no debe ser ignorada al evaluar el desempeño de estos dos jueces.

Debido a la paradoja de Simpson, es necesario tener mucho cuidado al sacar conclusiones cuando se usan datos agregados. Antes de cualquier conclusión acerca de la relación entre dos variables, en una tabulación cruzada en la que se usan datos agregados, es preciso investigar si no existen variables ocultas que afecten los resultados.

Diagrama de dispersión y línea de tendencia

Un **diagrama de dispersión** es una representación gráfica de la relación entre dos variables cuantitativas y una **línea de tendencia** es una línea que da una aproximación de la relación. Como ejemplo, considere la relación publicidad/ventas en una tienda de equipos de sonido. Durante los últimos tres meses, en 10 ocasiones la tienda apareció en comerciales de televisión, en el fin de semana, para promover sus ventas. Los directivos quieren investigar si hay relación entre el número de comerciales emitidos el fin de semana y las ventas en la semana siguiente. En la tabla 2.12 se presentan datos muestrales de las 10 semanas dando las ventas en cientos de dólares.

En la figura 2.7 aparece el diagrama de dispersión y la línea de tendencia* de los datos de la tabla 2.12. El número de comerciales (x) aparece en el eje horizontal y las ventas (y) en el eje vertical. En la semana 1, $x = 2$ y $y = 50$. En el diagrama de dispersión se grafica un punto con estas coordenadas. Para las otras nueve semanas se grafican puntos similares. Observe que en dos semanas sólo hubo un comercial, en otras dos semanas hubo dos comerciales, etcétera.

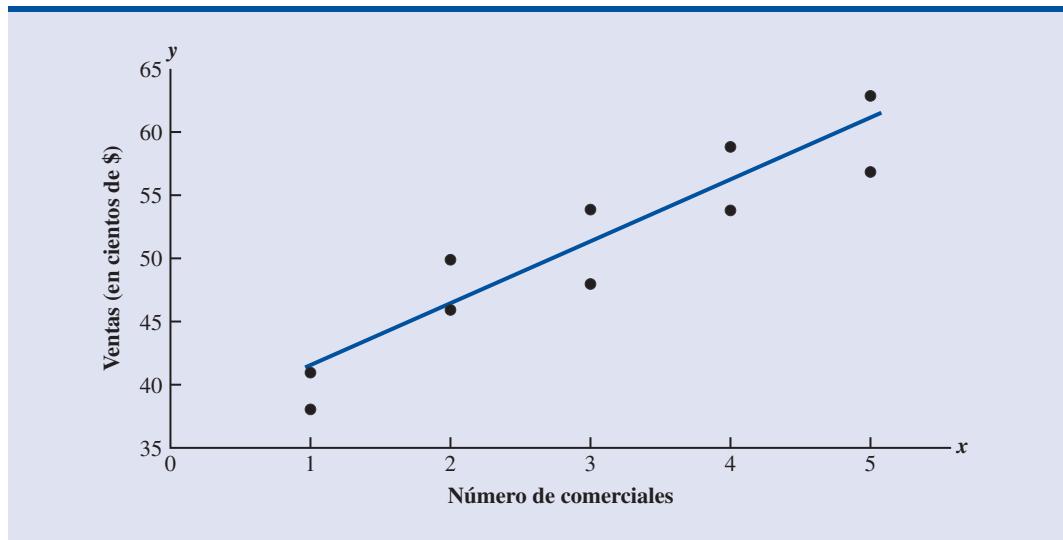
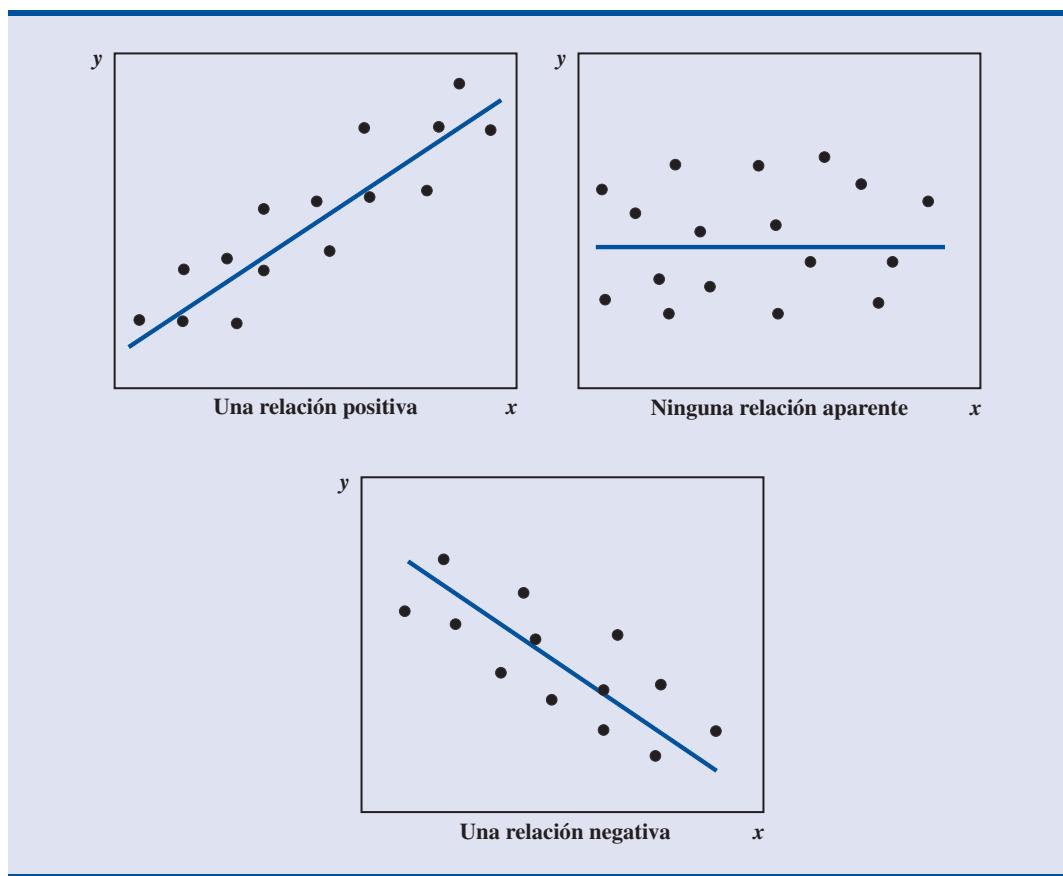
De nuevo, respecto de la figura 2.7, se observa una relación positiva entre el número de comerciales y las ventas. Más ventas corresponden a más comerciales. La relación no es perfecta ya que los puntos no trazan una línea recta. Sin embargo, el patrón que siguen los puntos y la línea de tendencia indican que la relación es positiva.

En la figura 2.8 se muestran los patrones de los diagramas de dispersión y el tipo de relación que sugieren. La gráfica arriba a la izquierda representa una relación positiva parecida a la del

TABLA 2.12 DATOS MUESTRALES DE UNA TIENDA DE EQUIPOS DE SONIDO

Semana	Número de comerciales		Ventas (en cientos de dólares)
	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	

*La ecuación de la línea de tendencia es $y = 36.15 + 4.95x$. La pendiente de la línea de tendencia es 4.95 y la intersección con el eje y (el punto en que la recta interseca el eje y) es 36.15. La interpretación de la pendiente y de la intersección con el eje y de una línea de tendencia lineal lo verá con detalle en el capítulo 12, cuando estudie la regresión lineal simple.

FIGURA 2.7 DIAGRAMA DE DISPERSIÓN Y LÍNEA DE TENDENCIA DE LA TIENDA DE EQUIPOS DE SONIDO**FIGURA 2.8** TIPOS DE RELACIÓN QUE APARECEN EN LOS DIAGRAMAS DE DISPERSIÓN

ejemplo de la cantidad de comerciales y las ventas. En la gráfica de arriba a la derecha no aparece ninguna relación entre las dos variables. La gráfica inferior representa una relación negativa en la que y tiende a disminuir a medida que x aumenta.

Ejercicios

Métodos



29. Los siguientes son datos de 30 observaciones en las que intervienen dos variables, x y y . Las categorías para x son A, B, y C; para y son 1 y 2.



Observación	x	y	Observación	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

- a. Con estos datos elabore una tabulación cruzada en la que x sea la variable para los renglones y y para las columnas.
 b. Calcule los porcentajes de los renglones.
 c. Calcule los porcentajes de las columnas.
 d. ¿Cuál es la relación, si hay alguna, entre las variables x y y ?

30. Las siguientes 20 observaciones corresponden a 20 variables cuantitativas, x y y .



Observación	x	y	Observación	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

- a. Elabore un diagrama de dispersión para la relación entre x y y .
 b. ¿Cuál es la relación, si hay alguna, entre x y y ?

Aplicaciones

31. En la siguiente tabulación cruzada se muestra el ingreso familiar de acuerdo con el nivel de estudios del jefe de familia, (*Statistical Abstract of the United States, 2002*).

Nivel de estudios	Ingreso por familia (en miles de dólares)					Total
	Menos de 25	25.0–49.9	50.0–74.9	75.0–99.9	100 o más	
No terminó secundaria	9 285	4 093	1 589	541	354	15 862
Terminó secundaria	10 150	9 821	6 050	2 737	2 028	30 786
Parte de bachillerato	6 011	8 221	5 813	3 215	3 120	26 380
Título universitario	2 138	3 985	3 952	2 698	4 748	17 521
Posgrado	813	1 497	1 815	1 589	3 765	9 479
Total	28 397	27 617	19 219	10 780	14 015	100 028

- a. Calcule los porcentajes por renglón e identifique las distribuciones de frecuencia porcentual del ingreso en los hogares en que el jefe de familia terminó secundaria y en los hogares en que el jefe de familia tiene un título universitario.
- b. ¿Qué porcentaje de las familias en que el jefe de familia terminó secundaria gana \$75 000 o más? ¿Qué porcentaje de las familias en que el jefe de familia tienen un título universitario gana 75 000 o más?
- c. Con los ingresos de los hogares en que el jefe de familia terminó secundaria elabore un histograma de la frecuencia porcentual, y otro con los ingresos de las familias en que el jefe de familia tiene un grado universitario. ¿Se observa alguna relación clara entre el ingreso familiar y el nivel de educación?
32. Consulte la tabulación cruzada del ingreso familiar de acuerdo con el nivel de estudios del ejercicio 31.
- a. Calcule los porcentajes e identifique las distribuciones de frecuencia porcentual. ¿Qué porcentaje de jefes de familia no terminó la secundaria?
- b. ¿Qué porcentaje de los hogares que perciben \$100 000 o más tienen como jefe de familia a una persona con un posgrado? ¿Qué porcentaje de los hogares que tienen como jefe de familia a una persona con un posgrado perciben más de \$100 000? ¿Por qué son diferentes estos dos porcentajes?
- c. Compare las distribuciones de frecuencia porcentual de aquellos hogares que perciben “Menos que 25”, “100 o más” y del “Total”. Haga un comentario sobre la relación entre ingreso familiar y nivel de estudios del jefe de familia.
33. Hace poco los administradores de un campo de golf recibieron algunas quejas acerca de las condiciones de los *greens*. Varios jugadores se quejaron de que estaban demasiado rápidos. En lugar de reaccionar a los comentarios de unos cuantos, la asociación de golf realizó un sondeo con 100 jugadoras y 100 jugadores. Los resultados del sondeo se presentan a continuación.

Jugadores		Condición de los greens		Jugadoras		Condición de los greens	
Hándicap		Demasiado rápido	Bien	Hándicap		Demasiado rápido	Bien
Menos de 15		10	40			1	9
15 o más		25	25			39	51

- a. Combine estas dos tabulaciones cruzadas utilizando como encabezados de renglón Jugadores y Jugadoras y como encabezados de columnas Demasiado rápido y Bien. ¿En qué grupo se encuentra el mayor porcentaje de los que dicen que los *greens* están demasiado rápidos?

- b. Vuelva a las tabulaciones cruzadas iniciales. De los jugadores con bajo hándicap (mejores jugadores), ¿en qué grupo (jugadoras o jugadores) se encuentra un porcentaje mayor de quienes dicen que los *greens* están demasiado rápidos?
- c. Regrese a las tabulaciones cruzadas iniciales. De los jugadores con alto hándicap, ¿en qué grupo (jugadoras o jugadores) se encuentra un porcentaje mayor para quienes los *greens* están demasiado rápidos?
- d. ¿Qué conclusiones obtiene acerca de mujeres y hombres respecto a la velocidad de los *greens*? ¿Las conclusiones que obtuvo en el inciso a son consistentes con los incisos b y c? Explique cualquier inconsistencia aparente.
34. En la tabla 2.13 se presentan datos financieros de 36 empresas de una muestra cuyas acciones cotizan en la bolsa de valores de Nueva York (*Investor's Business Daily*, 7 de abril de 2000). Los datos de la columna Ventas/margen/ROE son evaluaciones financieras compuestas que se basan en la tasa de crecimiento de las ventas de una empresa, su margen de ganancia y su rendimiento de los activos (ROE *return on capital employed*). La calificación EPS es una medida del crecimiento por acción.

TABLA 2.13 DATOS FINANCIEROS DE 36 EMPRESAS QUE CONFORMAN UNA MUESTRA

Empresa	EPS	Fuerza relativa del precio	Fuerza relativa del grupo de industrias	Ventas/margen/ ROE
Advo	81	74	B	A
Alaska Air Group	58	17	C	B
Alliant Tech	84	22	B	B
Atmos Energy	21	9	C	E
Bank of Am.	87	38	C	A
Bowater PLC	14	46	C	D
Callaway Golf	46	62	B	E
Central Parking	76	18	B	C
Dean Foods	84	7	B	C
Dole Food	70	54	E	C
Elec. Data Sys.	72	69	A	B
Fed. Dept. Store	79	21	D	B
Gateway	82	68	A	A
Goodyear	21	9	E	D
Hanson PLC	57	32	B	B
ICN Pharm.	76	56	A	D
Jefferson Plt.	80	38	D	C
Kroger	84	24	D	A
Mattel	18	20	E	D
McDermott	6	6	A	C
Monaco	97	21	D	A
Murphy Oil	80	62	B	B
Nordstrom	58	57	B	C
NYMAGIC	17	45	D	D
Office Depot	58	40	B	B
Payless Shoes	76	59	B	B
Praxair	62	32	C	B
Reebok	31	72	C	E
Safeway	91	61	D	A
Teco Energy	49	48	D	B
Texaco	80	31	D	C
US West	60	65	B	A
United Rental	98	12	C	A
Wachovia	69	36	E	B
Winnebago	83	49	D	A
York International	28	14	D	B

Fuente: *Investor's Business Daily*, 7 de abril de 2000.

- a. Elabore una tabulación cruzada con los datos Ventas/margen/ROE (renglones) y EPS (columnas). Para el EPS emplee las clases 0–19, 20–39, 40–59, 60–79 y 80–99.
 - b. Calcule los porcentajes de las columnas y haga un comentario sobre la relación entre las variables.
35. Regrese a la tabla 2.13.
- a. Elabore una tabulación cruzada con los datos Ventas/margen/ROE y Fuerza relativa del grupo de industrias.
 - b. Elabore una distribución de frecuencia de los datos Ventas/margen/ROE.
 - c. Elabore una distribución de frecuencia de los datos Fuerza relativa del grupo de industrias.
 - d. ¿Le ayudó la tabulación cruzada en la elaboración de las distribuciones de frecuencia de los incisos b y c?
36. De nuevo, a la tabla 2.13.
- a. Elabore un diagrama de dispersión con los datos EPS y Fuerza relativa del precio.
 - b. Haga un comentario sobre la relación entre las variables. (El significado del EPS se describe en el ejercicio 34. La Fuerza relativa del precio es una medida de la variación en el precio de una acción en los últimos 12 meses. Valores altos indican gran variación.)
37. La National Football League de Estados Unidos evalúa a los candidatos posición por posición con una escala que va de 5 a 9. La evaluación se interpreta como sigue: 8–9 debe empezar el primer año; 7.0–7.9 debe empezar; 6.0–6.9 será un apoyo para el equipo, y 5.0–5.9 puede pertenecer al club y contribuir. En la tabla 2.14 se presentan posición, peso, tiempo (segundos en correr 40 yardas), y evaluación de 40 candidatos (*USA Today*, 14 de abril de 2000).
- a. Con los datos posición (renglones) y tiempo (columnas) elabore una tabulación cruzada. Para el tiempo emplee las clases 4.00–4.49, 4.50–4.99, 5.00–5.49 y 5.50–5.99.
 - b. Haga un comentario acerca de la relación entre posición y tiempo, con base en la tabulación cruzada que elaboró en el inciso a.
 - c. Con los datos tiempo y calificación obtenida en la evaluación elabore un diagrama de dispersión, coloque la calificación obtenida en la evaluación en el eje vertical.
 - d. Haga un comentario sobre la relación entre tiempo y calificación obtenida en la evaluación.

Resumen

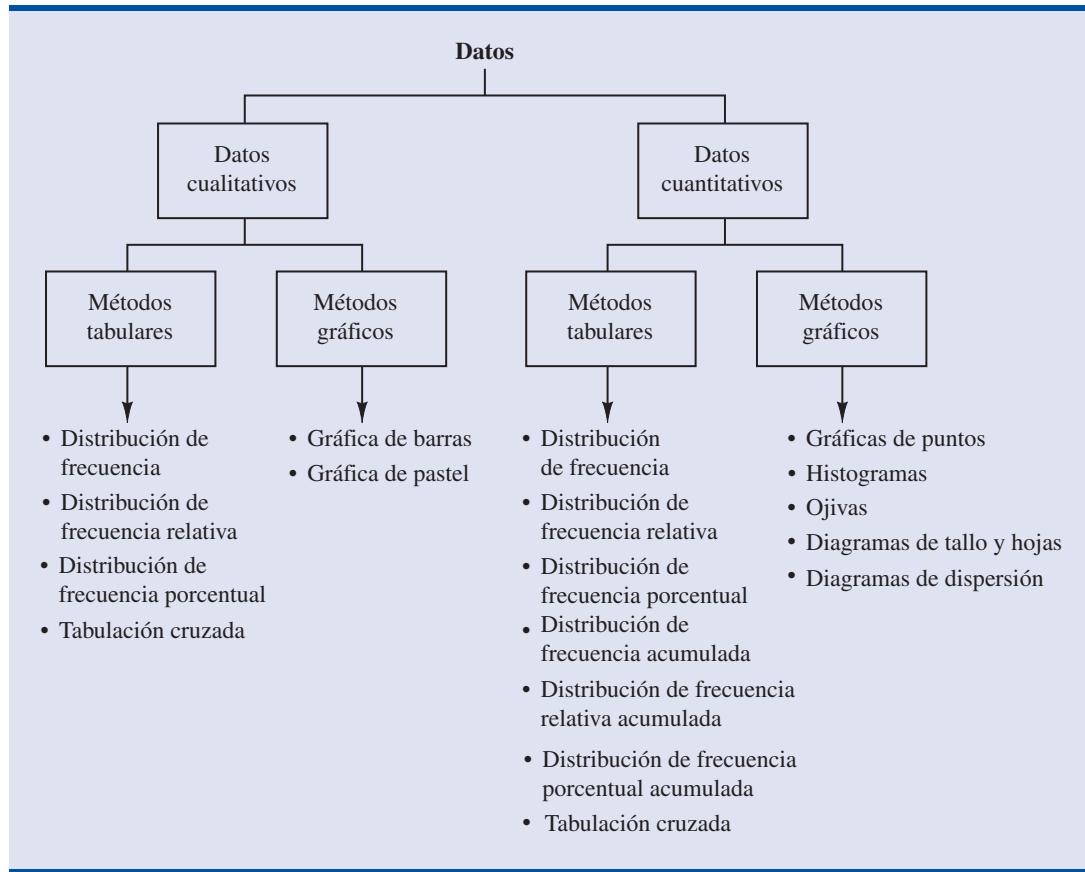
Un conjunto de datos, aunque sea de tamaño modesto, es difícil de interpretar con los datos tal y como se han recolectado. Los métodos tabulares y los métodos gráficos permiten organizar y resumir los datos para que muestren algún patrón y sean factibles de interpretación. Para resumir datos cualitativos se presentaron las distribuciones de frecuencia, de frecuencia relativa y las de frecuencia porcentual, las gráficas de barras y las gráficas de pastel. Las distribuciones de frecuencia, de frecuencia relativa, de frecuencia porcentual, los histogramas, las distribuciones de frecuencia acumulada, de frecuencia relativa acumulada, de frecuencia porcentual acumulada y las ojivas se presentaron como métodos para resumir datos cuantitativos. Los diagramas de tallo y hojas son una técnica para el análisis exploratorio de datos que se usa para resumir datos cuantitativos. La tabulación cruzada se presentó como un método para resumir datos para dos variables. Los diagramas de dispersión se presentaron como un método gráfico para mostrar la relación entre dos variables cuantitativas. En la figura 2.9 se resumen los métodos tabulares y gráficos que se presentaron en este capítulo.

Cuando se tienen grandes conjuntos de datos es indispensable usar paquetes de software para la elaboración de resúmenes tabulares o gráficos de los datos. En los dos apéndices de este capítulo se explica el uso de Minitab y de Excel con tal propósito.

TABLA 2.14 DATOS DE 40 CANDIDATOS A LA NATIONAL FOOTBALL LEAGUE DE ESTADOS UNIDOS



Observación	Nombre	Posición	Peso	Tiempo	Evaluación
1	Peter Warrick	Receptor abierto	194	4.53	9
2	Plaxico Burress	Receptor abierto	231	4.52	8.8
3	Sylvester Morris	Receptor abierto	216	4.59	8.3
4	Travis Taylor	Receptor abierto	199	4.36	8.1
5	Laveranues Coles	Receptor abierto	192	4.29	8
6	Dez White	Receptor abierto	218	4.49	7.9
7	Jerry Porter	Receptor abierto	221	4.55	7.4
8	Ron Dugans	Receptor abierto	206	4.47	7.1
9	Todd Pinkston	Receptor abierto	169	4.37	7
10	Dennis Northcutt	Receptor abierto	175	4.43	7
11	Anthony Lucas	Receptor abierto	194	4.51	6.9
12	Darrell Jackson	Receptor abierto	197	4.56	6.6
13	Danny Farmer	Receptor abierto	217	4.6	6.5
14	Sherrod Gideon	Receptor abierto	173	4.57	6.4
15	Trevor Gaylor	Receptor abierto	199	4.57	6.2
16	Cosey Coleman	Guardia	322	5.38	7.4
17	Travis Claridge	Guardia	303	5.18	7
18	Kaulana Noa	Guardia	317	5.34	6.8
19	Leander Jordan	Guardia	330	5.46	6.7
20	Chad Clifton	Guardia	334	5.18	6.3
21	Manula Savea	Guardia	308	5.32	6.1
22	Ryan Johanningmeir	Guardia	310	5.28	6
23	Mark Tauscher	Guardia	318	5.37	6
24	Blaine Saipaia	Guardia	321	5.25	6
25	Richard Mercier	Guardia	295	5.34	5.8
26	Damion McIntosh	Guardia	328	5.31	5.3
27	Jeno James	Guardia	320	5.64	5
28	Al Jackson	Guardia	304	5.2	5
29	Chris Samuels	Tacle ofensivo	325	4.95	8.5
30	Stockar McDougle	Tacle ofensivo	361	5.5	8
31	Chris McIngvish	Tacle ofensivo	315	5.39	7.8
32	Adrian Klemm	Tacle ofensivo	307	4.98	7.6
33	Todd Wade	Tacle ofensivo	326	5.2	7.3
34	Marvel Smith	Tacle ofensivo	320	5.36	7.1
35	Michael Thompson	Tacle ofensivo	287	5.05	6.8
36	Bobby Williams	Tacle ofensivo	332	5.26	6.8
37	Darnell Alford	Tacle ofensivo	334	5.55	6.4
38	Terrance Beadles	Tacle ofensivo	312	5.15	6.3
39	Tutan Reyes	Tacle ofensivo	299	5.35	6.1
40	Greg Robinson-Ran	Tacle ofensivo	333	5.59	6

FIGURA 2.9 MÉTODOS TABULARES Y GRÁFICOS PARA RESUMIR DATOS

Glosario

Datos cualitativos Etiquetas o nombres que se usan para identificar las categorías de elementos semejantes.

Datos cuantitativos Valores numéricos que indican cuánto o cuántos.

Distribución de frecuencia Resumen tabular de datos que muestra el número (frecuencia) de los datos que pertenecen a cada una de varias clases disyuntas.

Distribución de frecuencia relativa Resumen tabular de datos que muestra la proporción o la fracción de datos propios de cada una de varias clases disyuntas.

Distribución de frecuencia porcentual Resumen tabular de datos que muestra el porcentaje de datos que corresponden a cada una de varias clases disyuntas.

Gráfica de barras Gráfica para representar datos cualitativos que hayan sido resumidos en una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual.

Gráfica de pastel Gráfica para representar datos resumidos mediante una distribución de frecuencia relativa y que se basa en la subdivisión de un círculo en sectores que corresponden a la frecuencia relativa de las clases.

Punto medio de clase Valor que se encuentra a la mitad entre el límite de clase inferior y el límite de clase superior.

Gráfica de puntos Gráfica que resume datos mediante la cantidad de puntos sobre los valores de los datos que se encuentran en un eje horizontal.

Histograma Representación gráfica de una distribución de frecuencia, de frecuencia relativa o de frecuencia porcentual que se construye colocando los intervalos de clase sobre un eje horizontal y la frecuencia, la frecuencia relativa o la frecuencia porcentual sobre un eje vertical.

Distribución de frecuencia acumulada Síntesis tabular de datos cuantitativos, en la que se muestra el número de datos que son menores o iguales que el límite superior de cada clase.

Distribución de frecuencia relativa acumulada Resumen tabular de datos cuantitativos, en el que se muestra la proporción o fracción de datos que son menores o iguales que el límite superior de cada clase.

Distribución de frecuencia porcentual acumulada Síntesis tabular de datos cuantitativos, en la que se muestra el porcentaje de datos que son menores o iguales que el límite superior de cada clase.

Ojiva Gráfica de una distribución acumulada.

Análisis exploratorio de datos Métodos en los que se emplean cálculos aritméticos sencillos y gráficas fáciles de elaborar para resumir datos en forma rápida.

Diagrama de tallo y hojas Técnica para el análisis exploratorio de datos que tanto ordena por jerarquía datos cuantitativos como proporciona claridad acerca de la forma de la distribución.

Tabulación cruzada Resumen tabular de datos de dos variables. Las clases de una de las variables se representan como renglones; las clases de la otra variable como columnas.

Paradoja de Simpson Conclusiones que se obtienen de dos o más tabulaciones cruzadas y que se invierten cuando se agregan los datos en una sola tabulación cruzada.

Diagrama de dispersión Representación gráfica de la relación entre dos variables cuantitativas. A una variable se le asigna un eje horizontal y a la otra un eje vertical.

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

Fórmulas clave

Frecuencia relativa

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

Ancho aproximado de clase

$$\frac{\text{Dato mayor} - \text{Dato menor}}{\text{Número de clases}} \quad (2.2)$$

Ejercicios complementarios

38. Los cinco automóviles más vendidos en Estados Unidos durante 2003 fueron la camioneta Chevrolet Silverado/C/K, la camioneta Dodge Ram, la camioneta Ford F-Series, el Honda Accord y el Toyota Camry (*Motor Trend*, 2003). En la tabla 2.15 se presenta una muestra de 50 compras de automóviles.

TABLA 2.15 DATOS DE 50 COMPRAS DE AUTOMÓVILES

Silverado	Ram	Accord	Camry	Camry
Silverado	Silverado	Camry	Ram	F-Series
Ram	F-Series	Accord	Ram	Ram
Silverado	F-Series	F-Series	Silverado	Ram
Ram	Ram	Accord	Silverado	Camry
F-Series	Ram	Silverado	Accord	Silverado
Camry	F-Series	F-Series	F-Series	Silverado
F-Series	Silverado	F-Series	F-Series	Ram
Silverado	Silverado	Camry	Camry	F-Series
Silverado	F-Series	F-Series	Accord	Accord

- a. Elabore una distribución de frecuencia y otra de frecuencia porcentual.
 b. ¿Cuál es la camioneta y el automóvil de pasajeros más vendidos?
 c. Haga una gráfica de pastel.
39. El Higher Education Research Institute de UCLA cuenta con estadísticas sobre las áreas que son más elegidas por los estudiantes de nuevo ingreso. Las cinco más elegidas son arte y humanidades (A), administración de negocios (B), ingeniería (E), política (P) y ciencias sociales (S) (*The New York Times Almanac*, 2006). Otras áreas (O), entre las que se encuentran biología, física, ciencias de la computación y educación se agruparon todas en una sola categoría. Las siguientes fueron las áreas elegidas por 64 estudiantes de recién ingreso de una muestra.
- | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |
- a. Dé una distribución de frecuencia y otra de frecuencia porcentual.
 b. Elabore una gráfica de barras.
 c. ¿Qué porcentaje de los estudiantes de nuevo ingreso elige una de las cinco áreas más elegidas?
 d. ¿Cuál es el área más elegida por los estudiantes de nuevo ingreso? ¿Qué porcentaje de los estudiantes de nuevo ingreso elige esta área?
40. A los 100 mejores entrenadores de golf la revista *Golf Magazine* les preguntó, “¿Cuál es el aspecto más relevante que impide a los jugadores de golf desarrollar todo su potencial?” Las respuestas fueron falta de precisión, técnica de golpe inadecuada, actitud mental inadecuada, falta de energía, práctica insuficiente, tiro al hoyo inadecuado, juego corto inadecuado y estrategia de decisión inadecuada. A continuación se presentan los datos obtenidos (*Golf Magazine*, febrero de 2002):

Actitud mental	Actitud mental	Juego corto	Juego corto	Juego corto
Práctica	Precisión	Actitud mental	Precisión	Tiro al hoyo
Energía	Técnica de golpe	Precisión	Juego corto	Tiro al hoyo
Precisión	Actitud mental	Actitud mental	Precisión	Energía
Precisión	Precisión	Juego corto	Energía	Juego corto
Precisión	Tiro al hoyo	Actitud mental	Estrategia de decisión	Precisión
Juego corto	Energía	Actitud mental	Técnica de golpe	Juego corto
Práctica	Práctica	Actitud mental	Energía	Energía
Actitud mental	Juego corto	Actitud mental	Juego corto	Estrategia de decisión
Precisión	Juego corto	Precisión	Actitud mental	Juego corto
Actitud mental	Tiro al hoyo	Actitud mental	Actitud mental	Tiro al hoyo
Práctica	Tiro al hoyo	Práctica	Juego corto	Tiro al hoyo
Energía	Actitud mental	Juego corto	Práctica	Estrategia de decisión
Precisión	Juego corto	Precisión	Práctica	Tiro al hoyo
Precisión	Juego corto	Precisión	Juego corto	Tiro al hoyo
Precisión	Técnica de golpe	Juego corto	Actitud mental	Práctica
Juego corto	Juego corto	Estrategia de decisión	Juego corto	Juego corto
Práctica	Práctica	Juego corto	Práctica	Estrategia de decisión
Actitud mental	Estrategia de decisión	Estrategia de decisión	Energía	Juego corto
Precisión	Práctica	Práctica	Práctica	Precisión

- a. Elabore una distribución de frecuencia y otra de frecuencia porcentual.
 b. ¿Cuáles son los aspectos más relevantes que impiden a un jugador de golf desarrollar su potencial?
41. El rendimiento de dividendos son los beneficios anuales que paga una empresa, expresado como porcentaje del precio de una acción (Dividendo/precio de la acción × 100). En la tabla 2.16 se presenta el rendimiento de dividendos de las empresas del promedio industrial Dow Jones (*The Wall Street Journal*, 3 de marzo de 2006).
- a. Haga una distribución de frecuencia y una distribución de frecuencia porcentual.
 b. Haga un histograma.
 c. Aporte un comentario sobre la forma de la distribución.

TABLA 2.16 RENDIMIENTO DE DIVIDENDOS DE LAS EMPRESAS DEL PROMEDIO INDUSTRIAL DOW JONES.

archivo en CD
DivYield

Empresa	Rendimiento de dividendos	Empresa	Rendimiento de dividendos
AIG	0.9	Home Depot	1.4
Alcoa	2.0	Honeywell	2.2
Altria Group	4.5	IBM	1.0
American Express	0.9	Intel	2.0
AT&T	4.7	Johnson & Johnson	2.3
Boeing	1.6	JPMorgan Chase	3.3
Caterpillar	1.3	McDonald's	1.9
Citigroup	4.3	Merck	4.3
Coca-Cola	3.0	Microsoft	1.3
Disney	1.0	3M	2.5
DuPont	3.6	Pfizer	3.7
ExxonMobil	2.1	Procter & Gamble	1.9
General Electric	3.0	United Technologies	1.5
General Motors	5.2	Verizon	4.8
Hewlett-Packard	0.9	Wal-Mart Stores	1.3

- d. ¿Qué indican los resúmenes tabular y gráfico acerca de los dividendos de las empresas del promedio industrial Dow Jones?
- e. ¿Qué empresa tiene el más alto rendimiento de dividendos? Si hoy el precio de las acciones de esta empresa es \$20 por acción y usted compra 500 acciones, ¿cuál será el ingreso por dividendos que genere anualmente esta inversión?
42. Cada año en Estados Unidos, aproximadamente 1.5 millones de los estudiantes de educación superior presentan un examen de aptitud escolar (SAT, por sus siglas en inglés). Cerca de 80% de las universidades e instituciones de educación superior emplean las puntuaciones obtenidas por los estudiantes en este examen como criterio de admisión (*College Board*, marzo de 2006). A continuación se presentan las puntuaciones obtenidas en las áreas de matemáticas y expresión verbal por una muestra de estudiantes.
- | | | | | |
|------|------|------|------|------|
| 1025 | 1042 | 1195 | 880 | 945 |
| 1102 | 845 | 1095 | 936 | 790 |
| 1097 | 913 | 1245 | 1040 | 998 |
| 998 | 940 | 1043 | 1048 | 1130 |
| 1017 | 1140 | 1030 | 1171 | 1035 |
- a. Presente una distribución de frecuencia y un histograma de estas puntuaciones. La primera clase debe empezar en la puntuación 750 y la amplitud de clase deberá ser 100.
- b. Dé un comentario sobre la forma de la distribución.
- c. ¿Qué otras observaciones puede hacer acerca de estas puntuaciones con base en los resúmenes tabulares y gráficos?
43. La Asociación estadounidense de inversionistas independientes informa sobre 94 acciones fantasma. El término *fantasma* se refiere a que son acciones de empresas pequeñas o medianas que no son seguidas de cerca por las principales casas de bolsa. A continuación se presenta, de una muestra de 20 acciones fantasma, información sobre el lugar donde se comercializa la acción —bolsa

archivo en CD
SATScores

SATScores

Shadow

Acción	Bolsa de cambio	Ganancia por acción (\$)	Relación Precio/ganancia
Chemi-Trol	OTC	0.39	27.30
Candie's	OTC	0.07	36.20
TST/Impreso	OTC	0.65	12.70

(continúa)

Acción	Bolsa de cambio	Ganancia por acción	Relación precio/ganancia
Unimed Pharm.	OTC	0.12	59.30
Skyline Chili	AMEX	0.34	19.30
Cyanotech	OTC	0.22	29.30
Catalina Light.	NYSE	0.15	33.20
DDL Elect.	NYSE	0.10	10.20
Euphonix	OTC	0.09	49.70
Mesa Labs	OTC	0.37	14.40
RCM Tech.	OTC	0.47	18.60
Anuhco	AMEX	0.70	11.40
Hello Direct	OTC	0.23	21.10
Hilite Industries	OTC	0.61	7.80
Alpha Tech.	OTC	0.11	34.60
Wegener Group	OTC	0.16	24.50
U.S. Home & Garden	OTC	0.24	8.70
Chalone Wine	OTC	0.27	44.40
Eng. Support Sys.	OTC	0.89	16.70
Int. Remote Imaging	AMEX	0.86	4.70

- de Nueva York (NYSE), American Stock Exchange (AMEX) o directamente (OTC)— la ganancia por acción y la relación precio/ganancia.
- Con los datos de bolsa de cambio haga una distribución de frecuencia y otra de frecuencia relativa. ¿Cuál tiene más acciones fantasma?
 - Con los datos ganancia por acción y relación precio/ganancia elabore distribuciones de frecuencia y de frecuencia relativa. Para las ganancias por acción emplee las clases 0.00–0.19, 0.20–0.39, etc.; para la relación precio/ganancia use las clases 0.0–9.9, 10.0–19.9, etc. ¿Qué observaciones y comentarios puede hacer acerca de las acciones fantasma?
44. Los datos siguientes de la oficina de los censos de Estados Unidos proporcionan la población en millones de personas por estado (*The World Almanac*, 2006).

Estado	Población	Estado	Población	Estado	Población
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
Hawai	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		

- Elabore una distribución de frecuencia, una de frecuencia porcentual y un histograma. Use como ancho de clase 2.5 millones.
- Explique el sesgo de la distribución.
- ¿Qué observaciones puede hacer acerca de la población en los 50 estados?

45. *Drug Store News* (septiembre de 2002) proporciona datos sobre ventas de medicamentos de las principales farmacias de Estados Unidos. Los datos siguientes son ventas anuales en millones.

Farmacia	Ventas	Farmacia	Ventas
Ahold USA	\$ 1 700	Medicine Shoppe	\$ 1 757
CVS	12 700	Rite-Aid	8 637
Eckerd	7 739	Safeway	2 150
Kmart	1 863	Walgreens	11 660
Kroger	3 400	Wal-Mart	7 250

- a. Dé un diagrama de tallo y hojas.
 b. Indique cuáles son las ventas anuales menores, mayores e intermedias.
 c. ¿Cuáles son las dos farmacias mayores?
 46. A continuación se presentan las temperaturas diarias más altas y más bajas registradas en 20 ciudades de Estados Unidos (*USA Today*, 3 de marzo 2006).

Ciudad	Alta	Baja	Ciudad	Alta	Baja
Albuquerque	66	39	Los Angeles	60	46
Atlanta	61	35	Miami	84	65
Baltimore	42	26	Minneapolis	30	11
Charlotte	60	29	New 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. Con las temperaturas altas elabore un diagrama de tallo y hojas.
 b. Con las temperaturas bajas elabore un diagrama de tallo y hojas.
 c. Compare los dos diagramas y haga comentarios acerca de las diferencias entre las temperaturas más altas y las más bajas.
 d. Proporcione una distribución de frecuencia de las temperaturas más altas y de las más bajas.
 47. Vuelva al conjunto de datos sobre las temperaturas más altas y las temperaturas más bajas en 20 ciudades presentado en el ejercicio 46.
 a. Elabore un diagrama de dispersión que muestre la relación entre las dos variables, temperatura más alta y temperatura más baja.
 b. Aperte sus comentarios sobre la relación entre las temperaturas más elevadas y las más bajas.
 48. Se realizó un estudio sobre satisfacción en el empleo en cuatro ocupaciones. La satisfacción en el empleo se midió mediante un cuestionario de 18 puntos en el que a cada punto había que calificarlo con una escala del 1 al 5; las puntuaciones más altas correspondían a mayor satisfacción en el empleo. La suma de las calificaciones dadas a los 18 puntos proporcionaba una medida de

Ocupación	Satisfacción	Ocupación	Satisfacción	Ocupación	Satisfacción
Abogado	42	Terapeuta físico	78	Analista de sistemas	60
Terapeuta físico	86	Analista de sistemas	44	Terapeuta físico	59
Abogado	42	Analista de sistemas	71	Ebanista	78
Analista de sistemas	55	Abogado	50	Terapeuta físico	60

(continúa)

Ocupación	Satisfacción	Ocupación	Satisfacción	Ocupación	Satisfacción
Abogado	38	Abogado	48	Terapeuta físico	50
Ebanista	79	Ebanista	69	Ebanista	79
Abogado	44	Terapeuta físico	80	Analista de sistemas	62
Analista de sistemas	41	Analista de sistemas	64	Abogado	45
Terapeuta físico	55	Terapeuta físico	55	Ebanista	84
Analista de sistemas	66	Ebanista	64	Terapeuta físico	62
Abogado	53	Ebanista	59	Analista de sistemas	73
Ebanista	65	Ebanista	54	Ebanista	60
Abogado	74	Analista de sistemas	76	Abogado	64
Terapeuta físico	52				

- la satisfacción en el empleo de cada uno de los individuos de la muestra. Los datos obtenidos fueron los siguientes.
- Dé una tabulación cruzada para ocupación y satisfacción en el trabajo.
 - En la tabulación cruzada del inciso a calcule los porcentajes de renglones.
 - ¿Qué observaciones puede hacer respecto a la satisfacción en el trabajo en estas ocupaciones?
49. ¿Generan más ingresos las grandes empresas? Los datos siguientes muestran la cantidad de empleados y el ingreso anual de 20 de las empresas de *Fortune* 1000 (*Fortune*, 17 de abril de 2000).

Empresa	Empleados	Ingreso (en millones de \$)	Empresa	Empleados	Ingreso (en millones de \$)
Sprint	77 600	19 930	American Financial	9 400	3 334
Chase Manhattan	74 801	33 710	Fluor	53 561	12 417
Computer Sciences	50 000	7 660	Phillips Petroleum	15 900	13 852
Wells Fargo	89 355	21 795	Cardinal Health	36 000	25 034
Sunbeam	12 200	2 398	Borders Group	23 500	2 999
CBS	29 000	7 510	MCI Worldcom	77 000	37 120
Time Warner	69 722	27 333	Consolidated Edison	14 269	7 491
Steelcase	16 200	2 743	IBP	45 000	14 075
Georgia-Pacific	57 000	17 796	Super Value	50 000	17 421
Toro	1 275	4 673	H&R Block	4 200	1 669

- Haga un diagrama de dispersión para mostrar la relación entre las variables ingreso y empleados.
 - Haga un comentario sobre la relación entre estas variables.
50. En un sondeo realizado entre los edificios comerciales que son clientes de Cincinnati Gas & Electric Company se preguntaba cuál era el principal combustible que empleaban para la calefacción y en qué año se había construido el edificio. A continuación se presenta una parte del diagrama cruzado que se obtuvo con los datos.

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 esta tabulación cruzada dando los totales de los renglones y de las columnas.
- b. Dé las distribuciones de frecuencia de año de construcción y de tipo de combustible empleado.
- c. Haga una tabulación cruzada en la que se muestren los porcentajes de columnas.
- d. Elabore una tabulación cruzada en la que se muestren los porcentajes de renglones.
- e. Comente acerca de la relación entre año de construcción y tipo de combustible empleado.
51. La tabla 2.17 contiene parte de los datos que se encuentran en el archivo titulado Fortune en el disco compacto que viene con el libro. Este archivo proporciona fondos propios, valor de mercado y ganancias de las 50 empresas en una muestra de *Fortune 500*.

TABLA 2.17 DATOS EN UNA MUESTRA DE 50 EMPRESAS DE *FORTUNE 500*


Empresa	Fondos propios (en miles de \$)	Valor de mercado (en miles de \$)	Ganancias (en miles de \$)
AGCO	982.1	372.1	60.6
AMP	2 698.0	12 017.6	2.0
Apple Computer	1 642.0	4 605.0	309.0
Baxter International	2 839.0	21 743.0	315.0
Bergen Brunswick	629.1	2 787.5	3.1
Best Buy	557.7	10 376.5	94.5
Charles Schwab	1 429.0	35 340.6	348.5
.	.	.	.
.	.	.	.
Walgreen	2 849.0	30 324.7	511.0
Westvaco	2 246.4	2 225.6	132.0
Whirlpool	2 001.0	3 729.4	325.0
Xerox	5 544.0	35 603.7	395.0

- a. Con las variables fondos propios y ganancia elabore una tabulación cruzada. Para las ganancias emplee las clases 0–200, 200–400, ..., 1000–1200. Para fondos propios emplee las clases 0–1200, 1200–2400, ..., 4800–6000.
- b. En la tabulación cruzada del inciso a calcule los porcentajes de renglón.
- c. ¿Observa alguna relación entre ganancia y fondos propios?
52. Vuelva a la tabla 2.17.
- a. Con las variables valor de mercado y ganancia elabore una tabulación cruzada.
- b. En la tabulación cruzada del inciso a calcule los porcentajes de renglón.
- c. Haga un comentario sobre la relación entre las variables.
53. Vuelva a la tabla 2.17.
- a. Elabore un diagrama de dispersión que muestre la relación entre las variables ganancia y fondos propios.
- b. Haga un comentario sobre la relación entre las variables.
54. Vuelva a la tabla 2.17.
- a. Elabore un diagrama de dispersión que muestre la relación entre las variables valor de mercado y fondos propios.
- b. Haga un comentario sobre la relación entre las variables.

Caso problema 1 Las tiendas Pelican

Las tiendas Pelican, una división de National Clothing, es una cadena de tiendas de ropa para mujer que tiene sucursales por todo Estados Unidos. Hace poco la tienda realizó una promoción en la que envió cupones de descuento a todos los clientes de otras tiendas de National Clothing. Los datos obtenidos en una muestra de 100 pagos con tarjeta de crédito en las tiendas Pelican durante un día de la promoción se presentan en el archivo titulado PelicanStores. En la tabla 2.18 se mues-

TABLA 2.18 DATOS DE 100 COMPRAS CON TARJETA DE CRÉDITO REALIZADAS EN LAS TIENDAS PELICAN



Cliente	Tipo de cliente	Artículos	Ventas netas	Modo de pago	Género	Estado civil	Edad
1	Regular	1	39.50	Discover	Masculino	Casado	32
2	Promocional	1	102.40	Proprietary Card	Femenino	Casada	36
3	Regular	1	22.50	Proprietary Card	Femenino	Casada	32
4	Promocional	5	100.40	Proprietary Card	Femenino	Casada	28
5	Regular	2	54.00	MasterCard	Femenino	Casada	34
.
.
96	Regular	1	39.50	MasterCard	Femenino	Casada	44
97	Promocional	9	253.00	Proprietary Card	Femenino	Casada	30
98	Promocional	10	287.59	Proprietary Card	Femenino	Casada	52
99	Promocional	2	47.60	Proprietary Card	Femenino	Casada	30
100	Promocional	1	28.44	Proprietary Card	Femenino	Casada	44

tra parte de este conjunto de datos. El modo de pago Proprietary card se refiere a pagos realizados usando una tarjeta de crédito de National Clothing. A los clientes que hicieron compras usando un cupón de descuento se les denomina aquí promocionales y a quienes hicieron sus compras sin emplear cupón de descuento se les denomina regulares. Como a los clientes de las tiendas Pelican no se les enviaron cupones promocionales, los directivos consideran que las ventas hechas a quienes presentaron un cupón de descuento son ventas que de otro modo no se hubieran hecho. Es claro que Pelican espera que los clientes promocionales continúen comprando con ellos.

La mayor parte de las variables que aparecen en la tabla 2.18 se explican por sí mismas, pero dos de las variables deben ser aclaradas.

- Artículos El número total de artículos comprados
 Ventas netas Cantidad total cargada a la tarjeta de crédito

Los directivos de Pelican desean emplear estos datos muestrales para tener información acerca de sus clientes y para evaluar la promoción utilizando los cupones de descuento.

Informe para los directivos

Emplee los métodos tabulares y gráficos de la estadística descriptiva para ayudar a los directivos de Pelican a elaborar un perfil de sus clientes y a evaluar la promoción. Su informe debe contener, por lo menos, lo siguiente:

1. Distribuciones de frecuencia porcentual de las variables clave.
2. Una gráfica de barras o una gráfica de pastel que muestre el número de clientes correspondiente a cada modo de pago.
3. Una tabulación cruzada con el tipo de cliente (regular o promocional) frente a ventas netas. Haga un comentario sobre las semejanzas o diferencias que observe.
4. Un diagrama de dispersión para investigar la relación entre ventas netas y edad del cliente.

Caso problema 2 Industria cinematográfica

La industria cinematográfica es un negocio muy competitivo. En más de 50 estudios se producen de 300 a 400 películas por año y el éxito financiero de estas películas varía considerablemente. Las variables usuales para medir el éxito de una película son ventas brutas (en millones de \$) en el fin de semana del estreno, ventas brutas totales (en millones de \$), número de salas en que se presenta la película, semanas en las que la película se encuentra entre las 60 mejores en ventas

**TABLA 2.19** DATOS DEL ÉXITO DE 10 PELÍCULAS

Película	Ventas brutas en el estreno (en millones de \$)	Ventas brutas totales (en millones de \$)	Número de salas	Semanas en las 60 mejores
<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

brutas. Los datos de una muestra de 100 películas producidas en 2005 se encuentran en el archivo titulado Movies. La tabla 2.19 muestra los datos de las 10 primeras películas que se encuentran en este archivo.

Informe para los directivos

Emplee los métodos tabulares y gráficos de la estadística descriptiva para saber cómo contribuyen estas variables al éxito de una película. Su informe debe contener lo siguiente.

1. Resúmenes tabular y gráfico de las cuatro variables interpretando cada resumen acerca de la industria cinematográfica.
2. Un diagrama de dispersión para investigar la relación entre ventas brutas totales y ventas brutas en el fin de semana del estreno. Analícelo.
3. Un diagrama de dispersión para investigar la relación entre ventas brutas totales y número de salas. Analícelo.
4. Un diagrama de dispersión para investigar la relación entre ventas brutas totales y número de semanas entre las 60 mejores. Analícelo.

Apéndice 2.1 Uso de Minitab para presentaciones gráficas y tabulares

Minitab ofrece amplias posibilidades para la elaboración de resúmenes tabulares y gráficos de datos. Minitab se usa para elaborar diversos resúmenes gráficos y tabulaciones cruzadas. Los métodos gráficos son: gráfica de puntos, histograma, diagrama de tallo y hojas y diagrama de dispersión.

Gráficas de puntos

Para esta demostración emplee los datos de la tabla 2.4 sobre las duraciones de las auditorías. Los datos se encuentran en la columna C1 de la hoja de cálculo de Minitab. Con los pasos siguientes se generará una gráfica de puntos.

- Paso 1. Seleccionar el menú **Graph** y elegir **Dotplot**
- Paso 2. Seleccionar **One Y, Simple** y hacer clic en **OK**
- Paso 3. Cuando aparezca el cuadro de diálogo de Dotplot-One Y, Simple:

Ingresar C1 en el cuadro **Graph Variables**.

Hacer clic en **OK**





Histograma

Empleando los datos de la tabla 2.4 sobre las duraciones de las auditorías se explicará cómo se construye un histograma con las frecuencias sobre el eje vertical. Los datos están en la columna C1 de la hoja de cálculo de Minitab. Con los pasos siguientes se generará un histograma de las duraciones de las auditorías.

Paso 1. Seleccionar el menú **Graph**

Paso 2. Elegir **Histogram**

Paso 3. Seleccionar **Simple** y hacer clic en **OK**

Paso 4. Cuando aparezca el cuadro de diálogo **Histogram-Simple**:

 Ingresar C1 en el cuadro **Graph Variables**

 Hacer clic en **OK**

Paso 5. Cuando aparezca el histograma:

 Posicionar el cursor del mouse sobre cualquiera de las barras

 Dar doble clic

Paso 6. Cuando aparezca el cuadro de diálogo **Edit Bars**:

 Hacer clic en la pestaña **Binning**

 Seleccionar **Cutpoint** en Interval Type

 Seleccionar **Midpoint/Cutpoint positions** en Interval Definition

 Ingresar 10:35/5 en el cuadro **Midpoint/Cutpoint positions***

 Hacer clic en **OK**

Observe que Minitab también proporciona la posibilidad de mostrar los puntos medios de los rectángulos del histograma como escala en el eje *x*. Si se desea esta opción, se modifica el paso 6 seleccionando **Midpoint** en Interval Definition e ingresando 12:32/5 en el cuadro **Midpoint/Cutpoint positions**. Con estos pasos se obtiene el mismo histograma pero con los puntos medios, 12, 17, 22, 27 y 32, marcados en los rectángulos del histograma.

Diagrama de tallo y hojas



Se emplearán los datos de la tabla 2.8 sobre el examen de aptitudes para mostrar la construcción de un diagrama de tallo y hojas. Los datos se encuentran en la columna C1 de la hoja de cálculo de Minitab. Mediante los pasos siguientes se genera el diagrama extendido de tallo y hojas que se muestra en la sección 2.3.

Paso 1. Seleccionar el menú **Graph**

Paso 2. Elegir **Steam-and-Leaf**

Paso 3. Cuando aparezca el cuadro de diálogo **Steam-and-Leaf**:

 Ingresar C1 en el cuadro **Graph Variables**

 Hacer clic en **OK**

Diagrama de dispersión



Para demostrar la elaboración de un diagrama de dispersión se emplearán los datos de la tienda de equipos de sonido que se presentan en la tabla 2.12. Las semanas están numeradas del 1 al 10 en la columna C1, los datos del número de comerciales se encuentran en la columna C2 y los datos de las ventas están en la columna C3 de la hoja de cálculo de Minitab. Con los pasos siguientes se generará el diagrama de dispersión que se muestra en la figura 2.7.

*10:35/5 indica que 10 es el valor inicial del histograma, 35 es el valor final del histograma y 5 es el ancho de clase.

Paso 1. Seleccionar el menú **Graph**

Paso 2. Elegir **Scatterplot**

Paso 3. Elegir **Simple** y dar clic en **OK**

Paso 4. Cuando aparezca el cuadro de diálogo Scatterplot-Simple:

Ingresar C3 bajo **Y variables** y C2 bajo **X variables**.

Hacer clic en **OK**

Tabulación cruzada



Para demostrar la elaboración de una tabulación cruzada se usan los datos de *Zagat's Restaurant Review*, parte de los cuales se muestran en la tabla 2.9. Los restaurantes se encuentran numerados del 1 al 300 en la columna C1 de la hoja de cálculo de Minitab. Los datos sobre la calidad en la columna C2 y los precios en la columna C3.

Minitab sólo puede elaborar una tabulación cruzada con variables cualitativas y el precio es una variable cuantitativa. De manera que primero necesita codificar los precios especificando la clase a la que pertenece cada precio. Con los pasos siguientes se codificarán los precios haciendo cuatro clases de precios en la columna C4: \$10–19, \$20–29, \$30–39 y \$40–49.

Paso 1. Seleccionar el menú **Data**

Paso 2. Elegir **Code**

Paso 3. Elegir **Numeric to Text**

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

Ingresar C3 en el cuadro **Code data from columns**

Ingresar C4 en el cuadro **Into Columns**

Ingresar 10:19 en el primer cuadro **Original values** y \$10–19 en el cuadro adyacente **New**

Ingresar 20:29 en el primer cuadro **Original values** y \$20–29 en el cuadro adyacente **New**

Ingresar 30:39 en el primer cuadro **Original values** y \$30–39 en el cuadro adyacente **New**

Ingresar 40:49 en el primer cuadro **Original values** y \$40–49 en el cuadro adyacente **New**

Hacer clic en **OK**

Para cada precio de la columna C3 aparecerá ahora su categoría correspondiente en la columna C4. Ahora puede elaborar una tabulación cruzada para calidad y categoría de los precios usando los datos de las columnas C2 y C4. Con los pasos siguientes se creará una tabulación cruzada que contendrá la misma información que la tabla 2.10.

Paso 1. Seleccionar el menú **Stat**

Paso 2. Elegir **Tables**

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

Paso 4. Cuando aparezcan los cuadros: **Cross Tabulation** y **Chi-Square**:

Ingresar C2 en el cuadro **For rows** y C4 en el cuadro **For columns**

Seleccionar **Counts**

Hacer clic en **OK**

Apéndice 2.2 Uso de Excel para presentaciones gráficas y tabulares

Excel ofrece amplias posibilidades para la elaboración de resúmenes tabulares y gráficos de datos. En este capítulo se muestra cómo usar Excel para elaborar una distribución de frecuencia, gráficas de barras, gráficas de pastel, histogramas, tabulaciones cruzadas y diagramas de dispersión. Se presentan dos de las herramientas más potentes de Excel: el asistente para gráficos y el informe de tabla dinámica

Distribución de frecuencia y gráficas de barras con datos cualitativos

En esta sección se muestra el uso de Excel para la elaboración de una distribución de frecuencia y de una gráfica de barras con datos cualitativos. Ambas cosas se ilustran empleando los datos de la tabla 2.1 sobre ventas de refrescos.

Distribución de frecuencia Se empezará por mostrar el uso de la función COUNTIF para elaborar una distribución de frecuencia con los datos de la tabla 2.1. Consulte la figura 2.10 a medida que se presentan los pasos de esta explicación. La hoja de cálculo con las fórmulas (en la que se ven las funciones y fórmulas empleadas) aparece en segundo plano y la hoja de cálculo con los valores (en la que aparecen los resultados obtenidos con las funciones y fórmulas usadas) aparece en primer plano.

En las celdas A1:A51 se encuentra el título “Ventas de refrescos” y los datos de 50 ventas de refrescos. En las celdas C1:D1 también se ingresaron los títulos “Refresco” y “Frecuencia”. Los nombres de los cinco refrescos se ingresaron en las celdas C2:C6. Ahora se puede usar la función COUNTIF de Excel para contar cuántas veces aparece cada refresco en las celdas A2:A51. Para esto se siguen los pasos:

- Paso 1. Seleccionar la celda D2
- Paso 2. Ingresar =COUNTIF(\$A\$2:\$A\$51,C2)
- Paso 3. Copiar la celda D2 a las celdas D3:D6

FIGURA 2.10 DISTRIBUCIÓN DE FRECUENCIA DE LAS VENTAS DE REFRESCOS CONSTRUIDA EMPLEANDO LA FUNCIÓN COUNTIF DE EXCEL

	A	B	C	D	E
1	Ventas de refrescos		Refresco	Frecuencia	
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		A	B	C
9	Diet Coke		1 Ventas de refrescos		Refresco
10	Pepsi		2 Coke Classic		Frecuencia
45	Pepsi		3 Diet Coke		19
46	Pepsi		4 Pepsi		8
47	Pepsi		5 Diet Coke		5
48	Coke Classic		6 Coke Classic		13
49	Dr. Pepper		7 Coke Classic		5
50	Pepsi		8 Dr. Pepper		
51	Sprite		9 Diet Coke		
52			10 Pepsi		
			45 Pepsi		
			46 Pepsi		
			47 Pepsi		
			48 Coke Classic		
			49 Dr. Pepper		
			50 Pepsi		
			51 Sprite		
			52		

Nota: Los renglones 11-44 están ocultos.

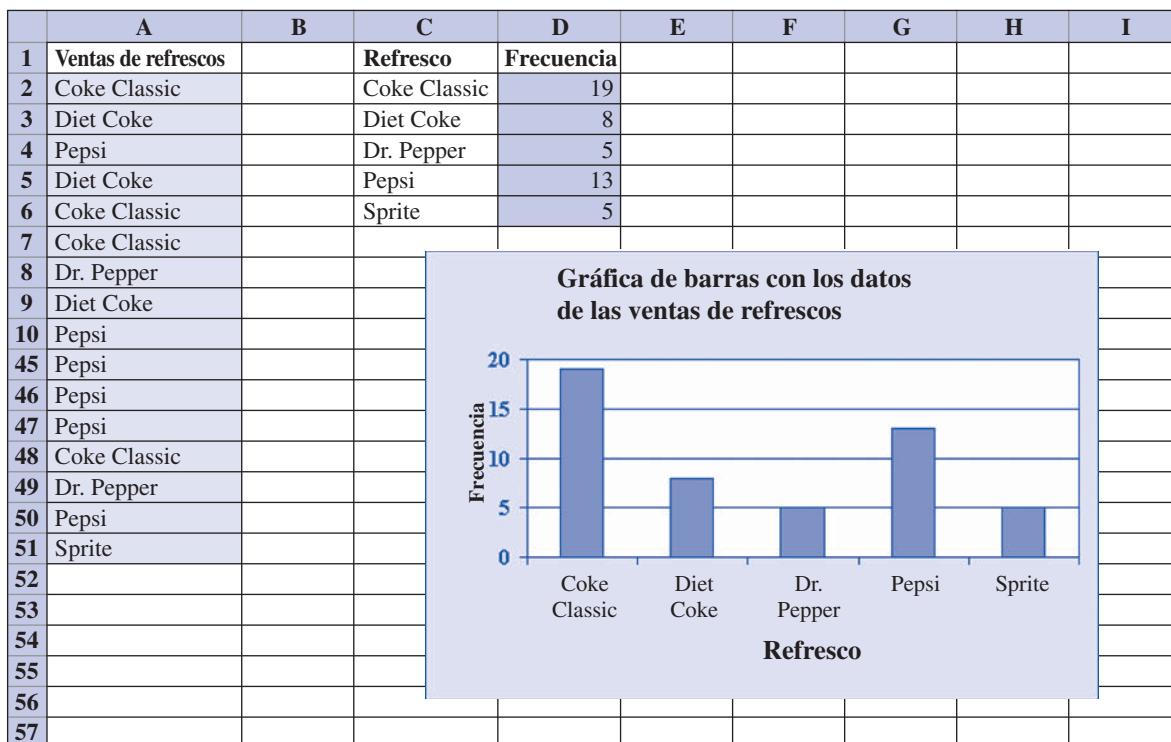
En la hoja de cálculo con las fórmulas de la figura 2.10 se observan en las celdas las fórmulas ingresadas al seguir estos pasos. En la hoja de cálculo con los valores se observan los valores obtenidos con las fórmulas de cada celda. En esta hoja de cálculo se aprecia la misma distribución de frecuencia de la tabla 2.2



Gráfica de barras Aquí se muestra cómo usar el asistente para gráficos de Excel para elaborar una gráfica de barras con los datos de las ventas de refrescos. En la figura 2.10 obsérvese la distribución de frecuencia que se presenta en la hoja de cálculo con los valores. La gráfica de barras que se va a construir es una extensión de esta hoja de cálculo. En la figura 2.11 se muestra esta misma hoja de cálculo con la gráfica de barras elaborada usando el asistente para gráficas. Los pasos a seguir son:

- Paso 1.** Seleccionar las celdas C1:D6
- Paso 2.** Hacer clic en el botón **Asistente para gráficas** de la barra de herramientas estándar (o seleccionar el menú **Insertar** y elegir la opción **Gráfico**)
- Paso 3.** Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 1 de 4: tipo de gráfico:
 Elegir **Columnas** de la lista **Tipo de gráfico**
 Elegir **Columnas agrupadas** en la visualización **Subtipo de gráfico**
 Hacer clic en **Siguiente >**
- Paso 4.** Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 2 de 4: datos de origen:
 Hacer clic en **Siguiente >**

FIGURA 2.11 GRÁFICA DE BARRAS CON LOS DATOS DE LAS VENTAS DE REFRESCOS ELABORADA MEDIANTE EL ASISTENTE PARA GRÁFICOS DE EXCEL



Paso 5. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 3 de 4: opciones de gráfico:

Seleccionar la pestaña **Títulos** y después

 Digitar Gráfica de barras con los datos de las ventas de refrescos
 en el cuadro **Título del gráfico**

 Digitar Refresco en el cuadro **Eje de categorías (X)**

 Digitar Frecuencia en el cuadro **Eje de valores (Y)**

Seleccionar la pestaña **Leyenda** y después

 Quitar la paloma (marca de verificación) que aparece en el cuadro
Mostrar leyenda

 Hacer clic en **Siguiente >**

Paso 6. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 4 de 4: ubicación:

 Especificiar una ubicación para la nueva gráfica (aquí se utilizó la misma hoja
 de cálculo que se estaba empleando por lo que se eligió la opción
Como objeto en)

 Hacer clic en **Finalizar**

En la figura 2.11 se muestra la gráfica de barras que se obtuvo.*

De manera similar, Excel puede elaborar una gráfica de pastel con los datos de las ventas de refrescos. La diferencia principal es que en el paso 3 se elige **Circular** de la lista Tipo de gráfico.

Distribuciones de frecuencia e histogramas para datos cuantitativos

En esta sección se muestra cómo usar Excel para elaborar una distribución de frecuencia y un histograma con datos cuantitativos. Para ilustrar esto se usan los datos de la tabla 2.4 sobre la duración de las auditorías.

Distribución de frecuencia Para elaborar una distribución de frecuencia con datos cuantitativos se puede usar la función FREQUENCY de Excel. Consulte la figura 2.12 a medida que se presentan los pasos a seguir. La hoja de cálculo con las fórmulas aparece en segundo plano y la hoja de cálculo con los valores aparece en primer plano. El título “Duración de la auditoría” se encuentra en la celda A1 y los datos de las 20 auditorías están en las celdas A2:A21. Siguiendo los procedimientos indicados en el texto, introduzca las cinco clases 10–14, 15–19, 20–24, 25–29 y 30–34. El título “Duración de la auditoría” y las cinco clases se ingresan en las celdas C1:C6. El título “Límite superior” y los cinco límites superiores de las clases se ingresan en las celdas D1:D6. Ingrese también el título “Frecuencia” en la celda E1. La función FREQUENCY de Excel se usará para obtener la frecuencia en las celdas E2:E6. Los pasos siguientes describen cómo elaborar una distribución de frecuencia con los datos de las duraciones de las auditorías.

Paso 1. Seleccionar las celdas E2:E6

Paso 2. Digitar, pero no ingresar, la fórmula siguiente:

=FREQUENCY(A2:A21,D2:D6)

Paso 3. Pulsar las teclas CTRL+SHIFT(mayúsculas)+ENTER con lo que la fórmula matricial será ingresada en cada una de las celdas E2:E6

El resultado se muestra en la figura 2.12. Los valores que aparecen en las celdas E2:E6 son las frecuencias de las clases correspondientes. Regrese a la función FREQUENCY, vea que el intervalo de las celdas para los límites superiores de clase (D2:D6) sirve de argumento a la función. Estos límites superiores de clase a los que Excel llama *bins*, le dicen a Excel qué frecuencia poner en las celdas del intervalo de salida (E2:E6). Por ejemplo, la frecuencia de la clase que tiene el límite superior, o *bin*, 14 será colocada en la primera celda (E2), la frecuencia de la clase que tiene el límite superior, o *bin*, 19 será colocada en la segunda celda (E3), y así sucesivamente.



Para ingresar una fórmula matricial es necesario mantener oprimidas las teclas Ctrl y Shift(mayúsculas) mientras se pulsa la tecla Enter.

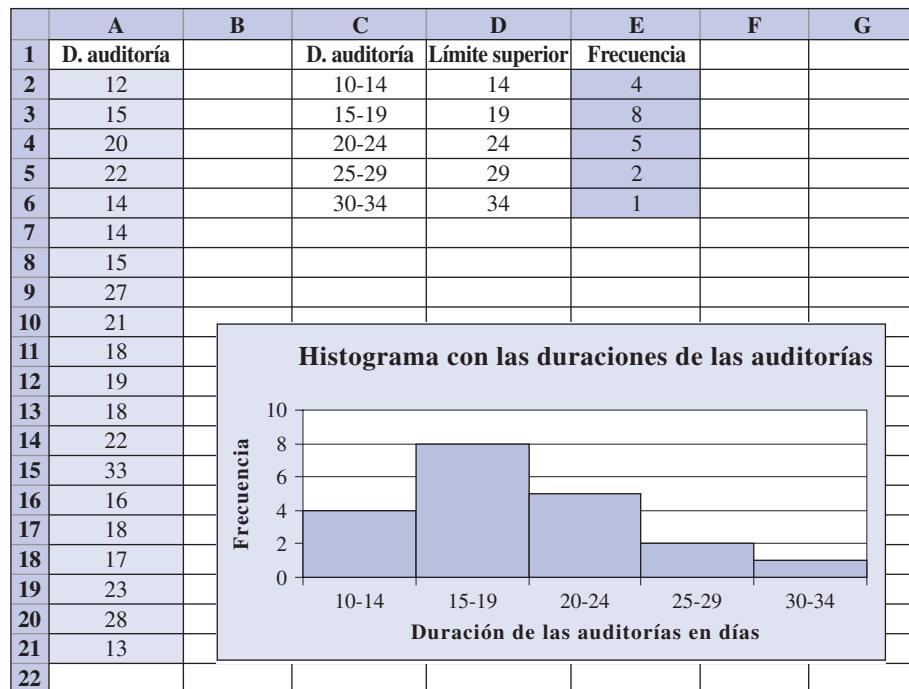
*La gráfica de barras de la figura 2.11 no es del mismo tamaño que la obtenida con Excel después de seleccionar **Finalizar**. Modificar el tamaño de una gráfica de Excel no es difícil. Primero se selecciona la gráfica, en los bordes de la gráfica aparecerán unos cuadritos negros llamados manillas de tamaño. Hacer clic sobre las manillas de tamaño y arrastrarlas para darle a la figura el tamaño deseado.

FIGURA 2.12 DISTRIBUCIÓN DE FRECUENCIA DE LOS DATOS DE LAS DURACIONES DE LAS AUDITORÍAS CON LA FUNCIÓN FREQUENCY DE EXCEL

	A	B	C	D	E
1	D. auditoría		D. auditoría	Límite superior	Frecuencia
2	12		10-14	14	=FREQUENCY(A2:A21,D2:D6)
3	15		15-19	19	=FREQUENCY(A2:A21,D2:D6)
4	20		20-24	24	=FREQUENCY(A2:A21,D2:D6)
5	22		25-29	29	=FREQUENCY(A2:A21,D2:D6)
6	14		30-34	34	=FREQUENCY(A2:A21,D2:D6)
7	14				
8	15		A	B	C
9	27		1	D. auditoría	D. auditoría
10	21		2	12	Límite superior
11	18		3	15	Frecuencia
12	19		4	20	
13	18		5	22	
14	22		6	14	
15	33		7	14	
16	16		8	15	
17	18		9	27	
18	17		10	21	
19	23		11	18	
20	28		12	19	
21	13		13	18	
			14	22	
			15	33	
			16	16	
			17	18	
			18	17	
			19	23	
			20	28	
			21	13	

Histograma Para usar el ayudante para gráficos de Excel para construir un histograma con las duraciones de las auditorías parte de la distribución de frecuencia de la figura 2.12. En la figura 2.13 se presenta la hoja de trabajo con la distribución de frecuencia y el histograma. Los pasos siguientes indican cómo emplear el asistente para gráficos al elaborar un histograma con los datos de las duraciones de las auditorías.

- Paso 1. Seleccionar las celdas E2:E6
- Paso 2. Hacer clic en el botón **Asistente para gráficas** de la barra de herramientas estándar (o seleccionar el menú **Insertar** y elegir la opción **Gráfico**)
- Paso 3. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 1 de 4: tipo de gráfico
 - Elegir **Columnas** en la lista **Tipo de gráfico**
 - Elegir **Columnas agrupadas** en la visualización **Subtipo de gráfico**
 - Hacer clic en **Siguiente >**
- Paso 4. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 2 de 4: datos de origen:
 - Seleccionar la pestaña **Serie** y después
 - Hacer clic en el cuadro **Rótulos del eje de categorías (X)**
 - Seleccionar las celdas C2:C6
 - Hacer clic en **Siguiente >**

FIGURA 2.13 HISTOGRAMA CON LAS DURACIONES DE LAS AUDITORÍAS

Paso 5. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 3 de 4: opciones de gráfico:

Seleccionar la pestaña **Títulos** y después

Digitalizar Histograma de las duraciones de las auditorías en el cuadro

Título del gráfico

Digitalizar Duración de las auditorías en días en el cuadro **Eje de categorías (X)**:

Digitalizar Frecuencia en el cuadro **Eje de valores (Y)**:

Seleccionar la pestaña **Leyenda** y después

Quitar la paloma (marca de verificación) que aparece en el cuadro
Mostrar leyenda

Hacer clic en **Siguiente >**

Paso 6. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 4 de 4: ubicación:

Especificar una ubicación para la nueva gráfica (aquí se utilizó la misma hoja de cálculo que se estaba empleando por lo que se eligió la opción

Como objeto en)

Hacer clic en **Finalizar**

Ahora en la hoja de cálculo aparecerá una gráfica de columnas elaborada por Excel. Pero entre las columnas habrá espacios. Como en un histograma no hay espacios entre las columnas, es necesario modificar esta gráfica para eliminar los espacios entre las columnas. Los pasos siguientes describen cómo hacerlo.

Paso 1. Dar doble clic en cualquiera de las columnas de la gráfica.

Paso 2. Cuando aparezca el cuadro de diálogo Formato de punto de datos:

Seleccionar la pestaña **Opciones**

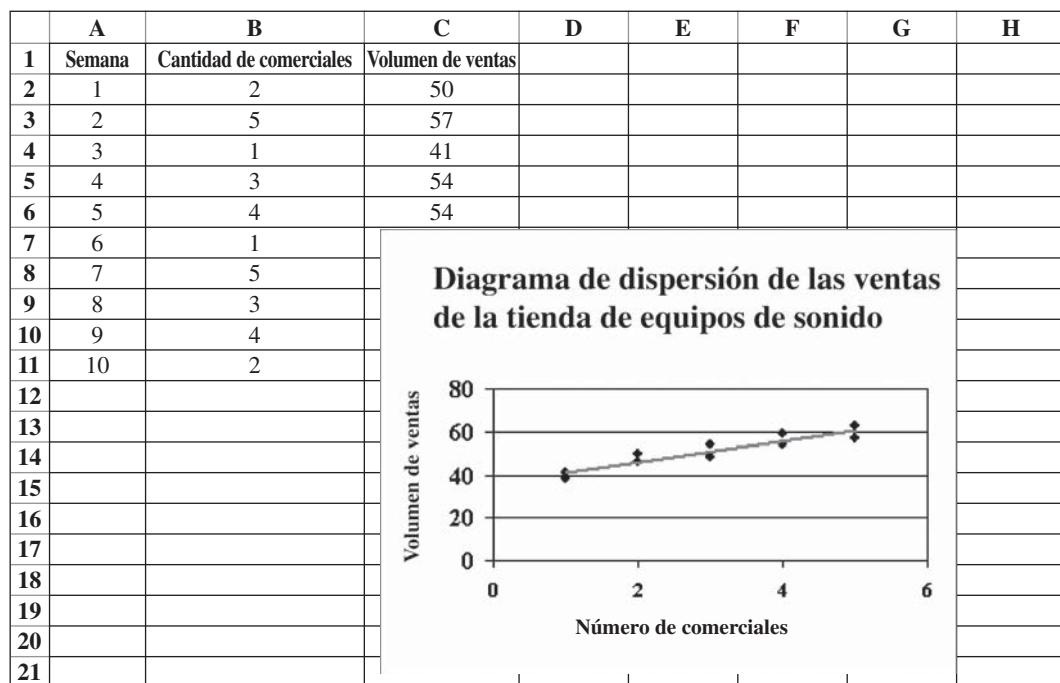
Ingresar 0 en el cuadro **Ancho del rango**

Hacer clic en **Aceptar**

El histograma se verá como el que aparece en la figura 2.13.

Por último, un aspecto interesante de la hoja de cálculo de la figura 2.13 es que Excel ha relacionado los datos que aparecen en las celdas A2:A21 con las frecuencias que aparecen en las celdas E2:E6 y con el histograma. Si se modifica alguno de los datos de las celdas A2:A21 se

FIGURA 2.14 DIAGRAMA DE DISPERSIÓN DE LAS VENTAS DE LA TIENDA DE EQUIPOS DE SONIDO



modificarán automáticamente las frecuencias de las celdas E2:E6 y también el histograma y aparecerán una distribución de frecuencias y un histograma modificados. Se aconseja probar cómo se realiza esta modificación automática modificando uno o dos de los datos.

Diagrama de dispersión



Se usarán los datos de la tienda de equipo de sonido que aparecen en la tabla 2.12 para mostrar cómo se usa el asistente para gráficos de Excel al elaborar un diagrama de dispersión. Consulte la figura 2.14 a medida que se describen los pasos para elaborar esta gráfica. La hoja de cálculo con los valores aparece en segundo plano y el diagrama de dispersión elaborado por el asistente para gráficos en primer plano. Los pasos a seguir son los siguientes.

- Paso 1. Seleccionar las celda B1:C11
- Paso 2. Hacer clic en el botón **Asistente para gráficos** de la barra de herramientas estándar (o seleccionar el menú **Insertar** y elegir la opción **Gráfico**)
- Paso 3. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 1 de 4: tipo de gráfico:
 - Elegir **XY (Dispersión)** en la lista **Tipo de gráfico**
 - Elegir **Dispersión** en la visualización **Subtipo de gráfico**
 - Hacer clic en **Siguiente >**
- Paso 4. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 2 de 4: datos de origen:
 - Hacer clic en **Siguiente >**
- Paso 5. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 3 de 4: opciones de gráfico:
 - Seleccionar la pestaña **Títulos**
 - Digitar **Diagrama de dispersión de las ventas de la tienda de equipos de sonido** en el cuadro **Título del gráfico**
 - Digitar **Número de comerciales** en el cuadro **Eje de categorías (X)**
 - Digitar **Volumen de ventas** en el cuadro **Eje de valores (Y)**:

Seleccionar la pestaña Leyenda

Quitar la paloma (marca de verificación) que aparece en el cuadro

Mostrar leyenda

Hacer clic en **Siguiente >**

Paso 6. Cuando aparezca el cuadro de diálogo Asistente para gráficos – paso 4 de 4: ubicación:

Especificar una ubicación para la nueva gráfica (aquí se utilizó la misma hoja de cálculo que se estaba empleando por lo que se eligió la opción

Como objeto en

Hacer clic en **Finalizar**

En el diagrama de dispersión puede trazar una línea de tendencia de la manera siguiente.

Paso 1. Colocar el cursor del mouse sobre cualquiera de los puntos del diagrama de dispersión y dar clic con el botón derecho del mouse. Aparecerá una lista de opciones

Paso 2. Elegir **Agregar Línea de tendencia**

Paso 3. Cuando aparezca el cuadro agregar línea de tendencia:

Seleccionar la pestaña **Tipo**

Elegir **Lineal** en la visualización **Tipo de tendencia o regresión**

Hacer clic en **Aceptar**

En la hoja de cálculo de la figura 2.14 se observa el diagrama de dispersión con la línea de tendencia.

Informe en tabla dinámica

El informe en tabla dinámica de Excel proporciona una valiosa herramienta para la manipulación de un conjunto de datos en que se tiene más de una variable. Se ilustrará su uso mostrando cómo elaborar una tabulación cruzada.

Tabulación cruzada Se ilustra la elaboración de una tabulación cruzada empleando los datos de los restaurantes que aparecen en la figura 2.15. Los títulos se han ingresado en el renglón 1 y los datos de los 300 restaurantes se han ingresado en las celdas A2:C301

FIGURA 2.15 HOJA DE CÁLCULO DE EXCEL CON LOS DATOS DE LOS RESTAURANTES

	A	B	C	D
1	Restaurante	Calidad	Precio (\$)	
2	1	Bueno	18	
3	2	Muy bueno	22	
4	3	Bueno	28	
5	4	Excelente	38	
6	5	Muy bueno	33	
7	6	Bueno	28	
8	7	Muy bueno	19	
9	8	Muy bueno	11	
10	9	Muy bueno	23	
11	10	Bueno	13	
292	291	Muy bueno	23	
293	292	Muy bueno	24	
294	293	Excelente	45	
295	294	Bueno	14	
296	295	Bueno	18	
297	296	Bueno	17	
298	297	Bueno	16	
299	298	Bueno	15	
300	299	Muy bueno	38	
301	300	Muy bueno	31	
302				

archivo
en
CD
Restaurant

Nota: los renglones
12–291 están
ocultos.

Paso 1. Seleccionar el menú **Datos**

Paso 2. Elegir **Informe de tabla y datos dinámicos**

Paso 3. Cuando aparezca el cuadro de diálogo Asistente para tablas y gráficos dinámicos – paso 1 de 3:

Elegir **Lista o base de datos de Microsoft Excel**

Elegir **Tabla dinámica**

Hacer clic en **Siguiente**

Paso 4. Cuando aparezca el cuadro de diálogo Asistente para tablas y gráficos dinámicos – paso 2 de 3:

Ingresar A1:C301 en el cuadro **Rango**

Hacer clic en **Siguiente**

Paso 5. Cuando aparezca el cuadro de diálogo Asistente para tablas y gráficos dinámicos – paso 3 de 3:

Seleccionar **Hoja de cálculo nueva**

Seleccionar **Diseño**

Paso 6. Cuando aparezca el diagrama Asistente para tablas y gráficos dinámicos – diseño (véase figura 2.16):

Arrastre el botón de campo **Calidad (Quality)** a la sección **FILA (ROW)** del diagrama

Arrastre el botón de campo **Precio (Meal Price)** a la sección **COLUMNA (COLUMN)** del diagrama

Arrastre el botón de campo **Restaurante (Restaurant)** a la sección **DATOS (DATA)** del diagrama

Dar doble clic en el botón de campo **Suma de Restaurante** en la sección DATOS
Cuando aparezca el cuadro de diálogo Campo de la tabla dinámica:

Elegir **Cuenta** bajo **Resumir por**

Hacer clic en **Aceptar** (la figura 2.17 muestra el diseño completo del diagrama)

Hacer clic en **Aceptar**

Paso 7. Cuando aparezca el cuadro de diálogo Asistente para tablas y gráficos dinámicos – paso 3 de 3:

Hacer clic en **Finalizar**

En la figura 2.18 se muestra parte del resultado generado por Excel. Observe que las columnas D a AK se han ocultado para que se puedan mostrar los resultados en una figura de tamaño razonable.

FIGURA 2.16 ASISTENTE PARA TABLAS Y GRÁFICOS DINÁMICOS: DISEÑO

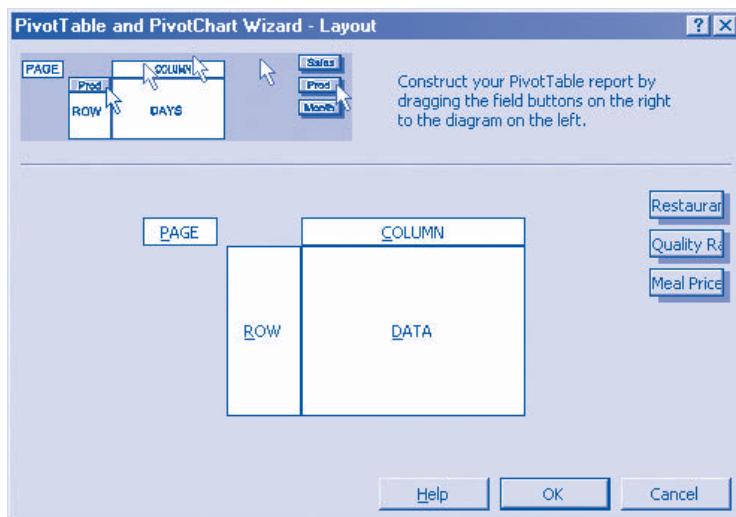
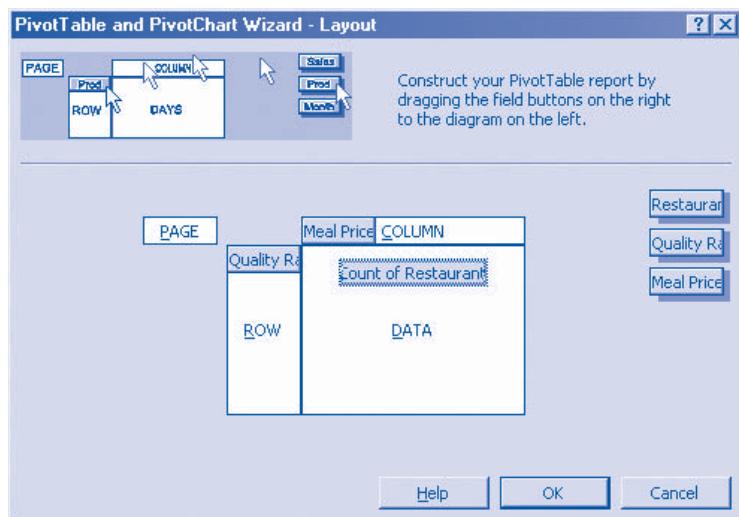


FIGURA 2.17 ASISTENTE PARA TABLAS Y GRÁFICOS DINÁMICOS: DISEÑO**FIGURA 2.18 RESULTADO INICIAL DEL INFORME DE TABLA DINÁMICA (LAS COLUMNAS D:AK ESTÁN OCULTAS)**

	A	B	C	AL	AM	AN	AO
1							
2							
3	Suma de restaurantes	Precio (\$)	■				
4	Calidad	■	10	11	47	48	Gran total
5	Excelente				2	2	66
6	Bueno		6	4			84
7	Muy bueno		1	4		1	150
8	Gran total		7	8	2	3	300
9							
10							
11							
12	<div style="border: 1px solid #ccc; padding: 5px; display: inline-block;"> PivotTable PivotTable ▾ </div>						
13							
14							
15							
16							
17							
18							
19							
20							

FIGURA 2.19 INFORME DE TABLA DINÁMICA FINAL CON LOS DATOS DE LOS RESTAURANTES

	A	B	C	D	E	F	G
1							
2							
3	Suma de restaurantes	Precio (\$)					
4	Calidad	■ 10-19	20-29	30-39	40-49	Gran total	
5	Bueno		42	40	2		84
6	Muy bueno		34	64	46	6	150
7	Excelente		2	14	28	22	66
8	Gran total		78	118	76	28	300
9							
10							
11							
12		PivotTable					
13		PivotTable ▾					
14							
15							
16							
17							
18							
19							
20							

nable. Los títulos de los renglones (Excelente, Bueno y Muy bueno) y los totales de los renglones (66, 84, 150 y 300) de la figura 2.18 son los mismos que en la tabla 2.10, sólo que en distinto orden. Para colocarlos en el orden Bueno, Muy bueno, Excelente hay que seguir los siguientes pasos:

Paso 1. Hacer clic con el botón derecho sobre la celda A5

Paso 2. Elegir **Ordenar**

Paso 3. Elegir **Mover al final**

En la figura 2.18 hay una columna para cada precio. Por ejemplo, en la columna B se encuentran los restaurantes cuyo precio es \$10, en la columna C los restaurantes cuyo precio es \$11, etc. Para que el informe en tabla dinámica se vea como en la tabla 2.10, se deben agrupar las columnas en cuatro categorías de precios: \$10–19, \$20–29, \$30–39 y \$40–49. Los pasos necesarios para agrupar las columnas de la hoja de cálculo que aparece en la figura 2.18 son:

Paso 1. Hacer clic con el botón derecho en Precio(\$) en la celda B3 de la Tabla dinámica

Paso 2. Elegir **Agrupar y mostrar detalles**

Elegir **Agrupar**

Paso 3. Cuando aparezca el cuadro de diálogo **Agrupar**

Ingresar 10 en el cuadro **Comenzar en**

Ingresar 49 en el cuadro **Terminar en**

Ingresar 10 en el cuadro **Por**

Hacer clic en **Aceptar**

La tabla dinámica que se obtiene se presenta en la figura 2.19. Es la tabla dinámica final. Obsérve que esta tabla proporciona la misma información que la tabla cruzada de la tabla 2.10.