

4

Objetivos de aprendizaje

Al concluir el capítulo, será capaz de:

OA1 Elaborar e interpretar un diagrama de puntos.

OA2 Crear e interpretar una gráfica de tallo y hojas.

OA3 Identificar y calcular medidas de posición.

OA4 Construir e interpretar diagramas de caja.

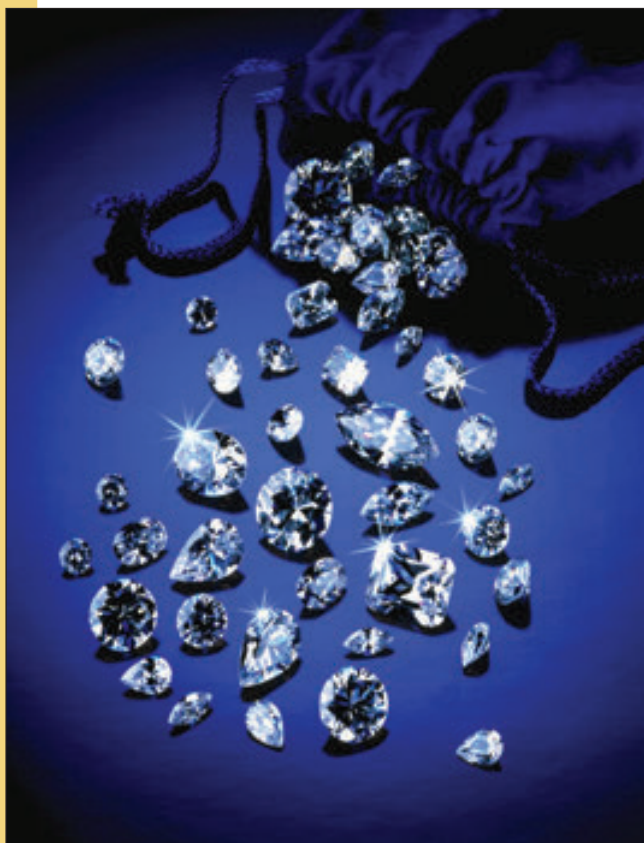
OA5 Calcular y entender el coeficiente de sesgo.

OA6 Trazar e interpretar un diagrama de dispersión.

OA7 Construir e interpretar una tabla de contingencia.

Descripción de datos:

Presentación y análisis de datos



Recientemente, McGivern Jewelers colocó un anuncio en el periódico local en el que informaba la forma, tamaño, precio y grado de corte de 33 de los diamantes que tenía en su inventario. Elabore el diagrama de caja de la variable *precio* y comente el resultado (vea ejercicio 37, objetivo 4).

4.1 Introducción

En el capítulo 2 se inició el estudio de la estadística descriptiva. Con el fin de transformar datos que están en bruto o no agrupados en alguna forma significativa, es necesario organizarlos en una distribución de frecuencias, la cual se representa en forma gráfica en un histograma o en un polígono de frecuencias. Este arreglo permite visualizar el lugar en donde tienden a acumularse los datos, los valores máximo y mínimo y la forma general de los datos.

En el capítulo 3 se calcularon primero diversas medidas de ubicación, tales como la media y la mediana, que permiten informar un valor típico de un conjunto de observaciones. También se calcularon diversas medidas de dispersión, tales como el rango y la desviación estándar, que permiten describir la variación o la dispersión en un conjunto de observaciones.

En este capítulo continúa el estudio de la estadística descriptiva. Se presentan los siguientes temas: 1) diagramas de puntos; 2) gráfica de tallo y hojas; 3) percentiles, y 4) diagramas de caja. Estos diagramas y la estadística proporcionan una idea adicional del lugar en el que los valores se concentran, así como de la forma general de los datos. En seguida se consideran datos bivariados de cada una de las observaciones individuales o seleccionadas. Algunos ejemplos incluyen: la cantidad de horas que estudia un alumno y los puntos que obtiene en un examen; si un producto tomado de la muestra es aceptable o no y el horario en el que se le fabrica; y la cantidad de electricidad que consume una casa en un mes, así como la temperatura alta media diaria de la región durante el mes.

4.2 Diagramas de puntos

OA1 Elaborar e interpretar un diagrama de puntos.

Los diagramas de dispersión dan una idea visual de la dispersión y concentración de los datos.

Recuerde que en los datos de Applewood Auto Group, la ganancia obtenida por la venta de 180 vehículos se resumió en ocho clases. Al organizar los datos en ocho clases se perdió el valor exacto de las observaciones. Por su parte, un **diagrama de puntos** agrupa los datos lo menos posible y evita la pérdida de identidad de cada observación. Para crear un diagrama de puntos se coloca un punto que representa a cada observación a lo largo de una recta numérica horizontal, la cual indica los valores posibles de los datos. Si hay observaciones idénticas o las observaciones se encuentran muy próximas, los puntos se *apilan* uno sobre otro para que se puedan ver de manera individual. Esto permite distinguir la forma de la distribución, el valor en torno al cual tienden a acumularse los datos y las observaciones máxima y mínima. Los diagramas de puntos son más útiles en el caso de conjuntos de datos pequeños, mientras que los histogramas lo son para conjuntos grandes de datos. Un ejemplo mostrará cómo construir e interpretar diagramas de puntos.

Ejemplo

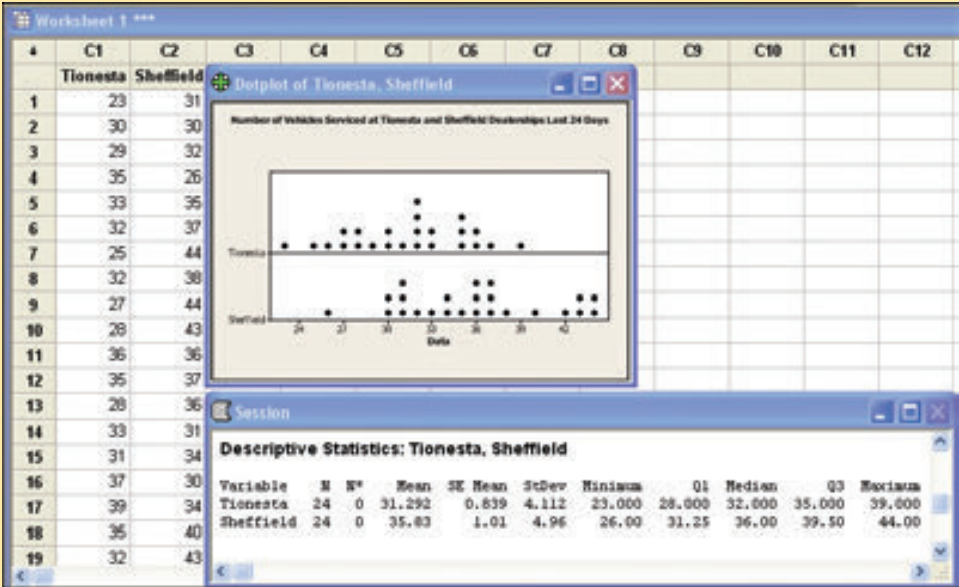
Los departamentos de servicio de Tionesta Ford Lincoln Mercury y Sheffield Motors Inc., dos de las cuatro distribuidoras de Applewood Auto Group, abrieron 24 días hábiles el mes pasado. A continuación aparece el número de vehículos que recibieron servicio el mes pasado en ambas distribuidoras. Elabore un diagrama de puntos y presente un resumen estadístico para comparar a estas dos distribuidoras.

Tionesta Ford Lincoln Mercury					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
23	33	27	28	39	26
30	32	28	33	35	32
29	25	36	31	32	27
35	32	35	37	36	30

Solución

Sheffield Motors Inc.					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
31	35	44	36	34	37
30	37	43	31	40	31
32	44	36	34	43	36
26	38	37	30	42	33

El sistema Minitab proporciona un diagrama de puntos y permite calcular la media, la mediana, los valores máximo y mínimo y la desviación estándar de la cantidad de automóviles que recibieron servicio en cada concesionaria durante los pasados 24 días hábiles.



Los esquemas de puntos que se muestran al centro de la captura de pantalla, ilustran gráficamente las distribuciones de ambas concesionarias. Los puntos muestran las diferencias en la ubicación y la dispersión de las observaciones. Al observar los esquemas de puntos, se puede ver que el número de vehículos que recibieron servicio en la distribuidora Sheffield están más dispersos y tienen una media mayor que los de Tionesta. Otras características del número de vehículos que recibieron servicio son:

- Tionesta dio servicio a menos vehículos en cualquier día dado, 23.
- Sheffield dio servicio a 26 autos en su día más bajo, 4 autos menos que en su siguiente día más bajo.
- Tionesta dio servicio exactamente a 32 vehículos en cuatro días diferentes.
- Los números de autos que recibieron servicio se acumulan alrededor del 36 en el caso de Sheffield y 32 en el de Tionesta.

A partir de la estadística descriptiva, es posible visualizar que Sheffield dio servicio a un promedio de 35.83 vehículos diarios y Tionesta, un promedio de 31.292 autos al día en el mismo periodo. También existe mayor dispersión, o variación, en el número diario de vehículos que recibieron servicio en Sheffield que en Tionesta. ¿Cómo se llega a esta conclusión? La desviación estándar de Sheffield es mayor (4.96 automóviles por día) que la de Tionesta (4.112 carros por día).

4.3 Gráficas de tallo y hojas

OA2 Crear e interpretar una gráfica de tallo y hojas.

En el capítulo 2 ilustramos la manera de organizar datos en una distribución de frecuencias de tal manera que permitiera resumir los datos brutos de forma significativa. La ventaja principal de organizar los datos en la distribución de frecuencias estriba en que nos permite visualizar de manera rápida la forma de la distribución sin necesidad de llevar a cabo ningún cálculo. En otras palabras, podemos ver dónde se concentran los datos y, asimismo, determinar si hay valores extremadamente grandes o pequeños. Sin embargo, hay dos desventajas que se presentan al organizar los datos en la distribución de frecuencias: 1) se pierde la identidad exacta de cada valor; 2) no es clara la forma en que los valores de cada clase se distribuyen. Para mayor precisión, la siguiente distribución de frecuencias muestra la cantidad de espacios publicitarios que compraron los 45 miembros de la Greater Buffalo Automobile Dealers Association durante el año 2010. Observe que 7 de las 45 concesionarias compraron de 90 a 100 espacios. Sin embargo, ¿los espacios comprados en esta clase se acumulan en torno a 90, se distribuyen uniformemente a lo largo de la clase o se acumulan cerca de 99? No es posible afirmar nada.

Cantidad de espacios comprados	Frecuencia
80 a 90	2
90 a 100	7
100 a 110	6
110 a 120	9
120 a 130	8
130 a 140	7
140 a 150	3
150 a 160	3
Total	45



Estadística en acción

En 1939 John W. Tukey (1915-2000) recibió un doctorado en matemáticas de Princeton. Sin embargo, cuando se unió a la Fire Control Research Office durante la Segunda Guerra Mundial, su interés en las matemáticas abstractas se desvió hacia la estadística aplicada. Ideó métodos numéricos y gráficos eficaces para estudiar los patrones que subyacían a los datos. Entre las gráficas que creó se encuentran el diagrama de tallo y hojas y el diagrama de caja y bigotes o diagrama de caja. De 1960 a 1980, Tukey encabezó la división de estadística electoral del equipo de proyección nocturno de la NBC. En 1960 se hizo famoso, ya que evitó el anuncio de la victoria anticipada de Richard Nixon en las elecciones presidenciales que ganó John F. Kennedy.

Otra técnica que se utiliza para representar información cuantitativa en forma condensada es el **diagrama de tallo y hojas**. Una ventaja de este diagrama sobre la distribución de frecuencias consiste en que no se pierde la identidad de cada observación. En el ejemplo anterior, no se conoce la identidad de los valores en la clase de 90 a 100. Para ilustrar la forma de construir un diagrama de tallo y hojas a partir de la cantidad de espacios publicitarios comprados, suponga que las siete observaciones en la clase del 90 a 100 son: 96, 94, 93, 94, 95, 96 y 97. El valor de **tallo** es el dígito o dígitos principales, en este caso 9. Las **hojas** son los dígitos secundarios. El tallo se coloca a la izquierda de una línea vertical y los valores de las hojas a la derecha.

Los valores en la clase de 90 a 100 se verían de la siguiente manera:

9		6	4	3	4	5	6	7
---	--	---	---	---	---	---	---	---

También es costumbre ordenar los valores en cada tallo de menor a mayor. Por consiguiente, la segunda fila del diagrama de tallo y hojas se vería de la siguiente manera:

9		3	4	4	5	6	6	7
---	--	---	---	---	---	---	---	---

Con un diagrama de tallo y hojas es más fácil observar que dos concesionarias compraron 94 espacios y que el número de espacios comprados varía de 93 a 97. Este tipo de diagrama se parece a una distribución de frecuencias, pero con mayor información, es decir, que la identidad de las observaciones se conserva.

DIAGRAMA DE TALLO Y HOJAS Técnica estadística para presentar un conjunto de datos. Cada valor numérico se divide en dos partes. El dígito principal se convierte en el tallo y los dígitos secundarios en las hojas. El tallo se localiza a lo largo del eje vertical y los valores de las hojas se apilan unos contra otros a lo largo del eje horizontal.

El siguiente ejemplo explica los detalles para elaborar un diagrama de tallo y hojas.

Ejemplo

La tabla 4-1 contiene la lista de la cantidad de espacios publicitarios de 30 segundos en radio que compró cada uno de los 45 miembros de la Greater Buffalo Automobile Dealers Association el año pasado. Organice los datos en un diagrama de tallo y hojas. ¿Alrededor de qué valores tiende a acumularse el número de espacios publicitarios? ¿Cuál es el número menor de espacios publicitarios comprados? ¿El número máximo de espacios comprados?

TABLA 4-1 Número de espacios publicitarios que compraron los miembros de la Greater Buffalo Automobile Dealers Association

96	93	88	117	127	95	113	96	108	94	148	156
139	142	94	107	125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139	134	119	97	89
118	136	125	143	120	103	113	124	138			

Solución

De acuerdo con los datos de la tabla 4-1, el número mínimo de espacios publicitarios comprados es de 88. Por ello, el primer valor de tallo es 8. El número máximo es 156, así que los valores de tallo comienzan en 8 y continúan hasta 15. El primer número de la tabla 4-1 es 96, que tendrá un valor de tallo de 9 y un valor de hoja de 6. Al desplazarnos por el renglón superior, el segundo valor es de 93 y el tercero de 88. Después de considerar los primeros tres valores de datos, el diagrama queda de la siguiente manera:

Tallo	Hoja
8	8
9	6 3
10	
11	
12	
13	
14	
15	

Al organizar los datos, el diagrama de tallo y hojas queda de la siguiente manera:

Tallo	Hoja
8	8 9
9	6 3 5 6 4 4 7
10	8 7 3 4 6 3
11	7 3 2 7 2 1 9 8 3
12	7 5 7 0 5 5 0 4
13	9 5 2 9 4 6 8
14	8 2 3
15	6 5 5

El procedimiento acostumbrado consiste en ordenar los valores de las hojas de menor a mayor. La última línea, la fila que se refiere a los valores próximos a 150, se vería de la siguiente manera:

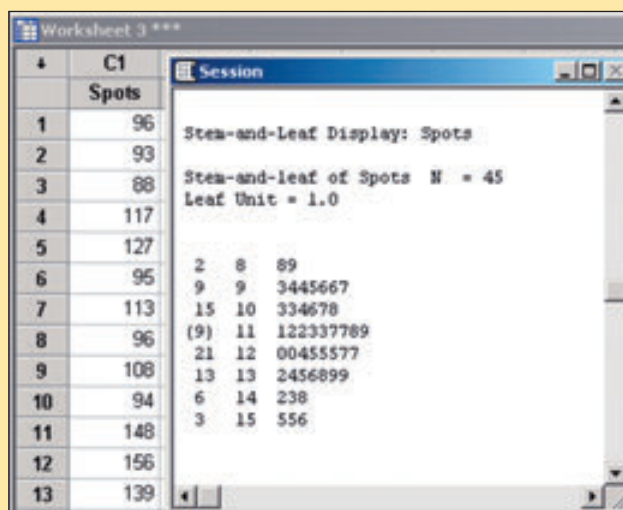
15		5	5	6
----	--	---	---	---

La tabla final sería la siguiente, en la cual están ordenados todos los valores de las hojas:

Tallo	Hoja
8	8 9
9	3 4 4 5 6 6 7
10	3 3 4 6 7 8
11	1 2 2 3 3 7 7 8 9
12	0 0 4 5 5 5 7 7
13	2 4 5 6 8 9 9
14	2 3 8
15	5 5 6

Es posible deducir algunas conclusiones del diagrama de tallo y hojas. Primero, la cantidad mínima de espacios publicitarios comprados es de 88 y la máxima de 156. Dos concesionarias compraron menos de 90 espacios, y tres compraron 150 o más. Observe, por ejemplo, que las tres concesionarias que compraron más de 150 espacios, en realidad compraron 155, 155 y 156 espacios. La concentración de la cantidad de espacios se encuentra entre 110 y 130. Hubo nueve concesionarias que compraron entre 110 y 119 espacios y ocho compraron entre 120 y 129 espacios. También note que en el grupo ubicado entre 120 y 129 el número real de espacios comprados se distribuyó uniformemente. Es decir, que dos concesionarias compraron 120 espacios, una compró 124 espacios, tres compraron 125 espacios y dos compraron 127 espacios.

Además, es posible generar esta información en el sistema de software Minitab. La variable se llama *Spots*. Abajo aparece la captura de pantalla de Minitab. Al final del capítulo usted puede encontrar los comandos de Minitab, que generan esta salida.



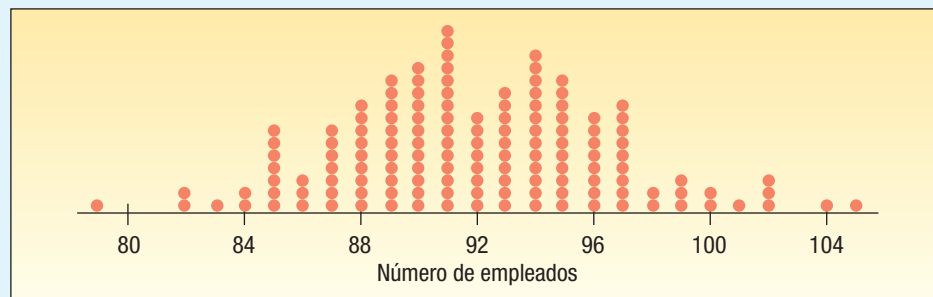
La solución de Minitab proporciona información adicional relacionada con los totales acumulados. En la columna a la izquierda de los valores de tallo se encuentran números como 2, 9, 15, y así sucesivamente. El número 9 indica que se presentaron 9 observaciones antes del valor de 100. El 15 muestra que se presentaron 15 observaciones antes de 110. Más o menos a la mitad de la columna aparece el número 9 entre paréntesis, que indica que el valor de en medio o mediana aparece en dicha fila y que hay nueve valores en este grupo. En este caso, el valor medio es el valor debajo del cual se presenta la mitad de las observaciones. Hay un total de 45 observaciones, así que el valor medio, en caso de que los datos se ordenen de menor a mayor, sería la observación vigésima tercera; este valor es 118. Después de la mediana, los valores comienzan a decrecer. Estos valores representan los totales acumulados *más que*. Hay 21 observaciones de 120 o más, 13 de 130 o más, y así sucesivamente.

¿Cuál es mejor: el esquema de puntos o el diagrama de tallo y hojas? En realidad, este dilema es cuestión de elección y conveniencia personal. Para presentar datos, en especial con una gran cantidad de observaciones, usted se dará cuenta de que los diagramas de puntos se utilizan con mayor frecuencia. Encontrará diagramas de puntos en la literatura analítica, informes de marketing y, en ocasiones, informes anuales. Si realiza un análisis rápido para usted mismo, los diagramas de tallo y hojas son accesibles y fáciles, en particular en relación con un conjunto pequeño de datos.

Autoevaluación 4-1



1. El siguiente diagrama muestra el número de empleados en cada una de las 142 tiendas de Home Depot ubicadas al sureste de Estados Unidos.



- a) ¿Cuáles son los números máximo y mínimo de empleados por tienda?
 - b) ¿Cuántas tiendas emplean a 91 personas?
 - c) ¿Alrededor de qué valores tiende a acumularse el número de empleados por tienda?
2. La tasa de recuperación de 21 acciones es la siguiente:

8.3	9.6	9.5	9.1	8.8	11.2	7.7	10.1	9.9	10.8
10.2	8.0	8.4	8.1	11.6	9.6	8.8	8.0	10.4	9.8
									9.2

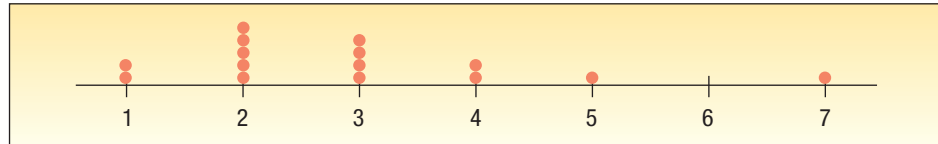
Organice esta información en un diagrama de tallo y hojas.

- a) ¿Cuántas tasas son menores que 9.0?
- b) Haga una lista de las tasas en la categoría que va de 10.0 a 11.0.
- c) ¿Cuál es la mediana?
- d) ¿Cuáles son las tasas máxima y mínima de recuperación?

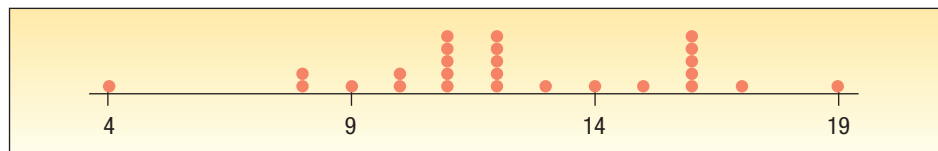
Ejercicios



1. Describa las diferencias entre un histograma y un diagrama de puntos. ¿Cuándo podría resultar mejor un diagrama de puntos que un histograma?
2. Explique las diferencias entre un histograma y un diagrama de tallo y hojas.
3. Considere el siguiente diagrama.



- a) ¿Qué nombre recibe este diagrama?
 - b) ¿Cuántas observaciones hay en el estudio?
 - c) ¿Cuáles son los valores máximo y mínimo?
 - d) ¿En torno a qué valores tienden a acumularse las observaciones?
4. El siguiente diagrama informa el número de teléfonos celulares que vendió Radio Shack durante los pasados 26 días.



- a) ¿Cuáles son los números máximo y mínimo de teléfonos celulares vendidos en un día?
 - b) ¿Cuál es el número típico de teléfonos celulares vendidos?
5. La primera fila del diagrama de tallo y hojas es la siguiente: 62 | 1 3 3 7 9. Suponga que se trata de números enteros.
 - a) ¿Cuál es el *posible rango* de los valores de esta fila?
 - b) ¿Cuántos valores de datos hay en esta fila?
 - c) Haga una lista de los valores reales de esta fila de datos.
 6. La tercera fila de un diagrama de tallo y hojas aparece de la siguiente manera: 21 | 0 1 3 5 7 9. Suponga que los valores son números enteros.
 - a) ¿Cuál es el *posible rango* de los valores de esta fila?
 - b) ¿Cuántos valores de datos hay en esta fila?
 - c) Elabore una lista de los valores reales de esta fila de datos.
 7. El siguiente diagrama de tallo y hojas del software de Minitab muestra el número de unidades producidas por día en una fábrica.

1	3	8
1	4	
2	5	6
9	6	0133559
(7)	7	0236778
9	8	59
7	9	00156
2	10	36

- a) ¿Cuántos días se registraron?
- b) ¿Cuántas observaciones hay en la primera clase?
- c) ¿Cuál es el valor mínimo y el valor máximo?

- d) Elabore una lista de los valores reales de la cuarta fila.
 e) Elabore una lista de los valores reales de la segunda fila.
 f) ¿Cuántos valores son menores que 70?
 g) ¿Cuántos valores son iguales a 80 o más?
 h) ¿Cuál es la mediana?
 i) ¿Cuántos valores se encuentran entre 60 y 89, inclusive?
8. El siguiente diagrama de tallo y hojas presenta la cantidad de películas rentadas por día en Video Connection, ubicado en la esquina de las calles Forth y Main.

3	12	689
6	13	123
10	14	6889
13	15	589
15	16	35
20	17	24568
23	18	268
(5)	19	13456
22	20	034679
16	21	2239
12	22	789
9	23	00179
4	24	8
3	25	13
1	26	
1	27	0

- a) ¿Cuántos días se registraron?
 b) ¿Cuántas observaciones hay en la última clase?
 c) ¿Cuáles son los valores máximo y mínimo de todo el conjunto de datos?
 d) Elabore una lista de valores reales de la cuarta fila.
 e) Elabore una lista de valores reales que aparecen en la penúltima fila.
 f) ¿En cuántos días se rentaron menos de 160 películas?
 g) ¿En cuántos días se rentaron 220 o más películas?
 h) ¿Cuál es el valor medio?
 i) ¿En cuántos días se rentaron entre 170 y 210 películas?
9. Una encuesta sobre el número de llamadas telefónicas por celular realizada con una muestra de suscriptores de Verizon la semana pasada reveló la siguiente información. Elabore un diagrama de tallo y hojas. ¿Cuántas llamadas hizo un suscriptor típico? ¿Cuáles fueron los números máximo y mínimo de llamadas que realizaron?

52	43	30	38	30	42	12	46	39
37	34	46	32	18	41	5		

10. Aloha Banking Co. estudia el uso de cajeros automáticos en los suburbios de Honolulu. Una muestra de 30 cajeros mostró que éstos se utilizaron la siguiente cantidad de veces el día de ayer. Elabore un diagrama de tallo y hojas. Resuma la cantidad de veces que se utilizó cada cajero automático. ¿Cuáles son los números mínimo y máximo de veces que se utilizó cada uno de ellos?

83	64	84	76	84	54	75	59	70	61
63	80	84	73	68	52	65	90	52	77
95	36	78	61	59	84	95	47	87	60

4.4 Otras medidas de posición

OA3 Identificar y calcular medidas de posición.

Los cuartiles dividen un grupo de datos en cuatro partes.

La desviación estándar es la medida de dispersión que más se utiliza. No obstante, existen otras formas de describir la variación o dispersión de un conjunto de datos. Un método consiste en determinar la *ubicación* de los valores que dividen un conjunto de observaciones en partes iguales. Estas medidas incluyen los **cuartiles**, **deciles** y **percentiles**.

Los cuartiles dividen a un conjunto de observaciones en cuatro partes iguales. Para explicarlo mejor, piense en un conjunto de valores ordenados de menor a mayor. En el capítulo 3 denominamos *mediana* al valor intermedio de un conjunto de datos ordenados de menor a mayor. Es decir que 50% de las observaciones son mayores que la mediana y 50% son menores. La mediana constituye una medida de ubicación, ya que señala el centro de los datos. De igual manera, los **cuartiles** dividen a un conjunto de observaciones en cuatro partes iguales. El primer cuartil, que se representa mediante Q_1 , es el valor debajo del cual se presenta 25% de las observaciones, y el tercer cuartil, que simboliza Q_3 , es el valor debajo del cual se presenta 75% de las observaciones. Lógicamente, Q_2 es la mediana. Q_1 puede considerarse como la *mediana* de la mitad inferior de los datos y Q_3 como la *mediana* de la parte superior de los datos.

Asimismo, los **deciles** dividen un conjunto de observaciones en 10 partes iguales y los **percentiles** en 100 partes iguales. Por lo tanto, si su promedio general en la universidad se encuentra en el octavo decil, usted podría concluir que 80% de los estudiantes tuvieron un promedio general inferior al suyo y 20%, un promedio superior. Un promedio general ubicado en el trigésimo tercer percentil significa que 33% de los estudiantes tienen un promedio general más bajo y 67% un promedio general más alto. Con frecuencia, en Estados Unidos, las calificaciones que se expresan en percentiles se utilizan para dar a conocer resultados relacionados con pruebas estandarizadas como SAT, ACT, GMAT (que se emplean para determinar el ingreso en algunas maestrías de administración de empresas) y LSAT (que sirve para determinar el ingreso a la escuela de leyes).

Cuartiles, deciles y percentiles

Para formalizar el proceso de cálculo, suponga que L_p representa la ubicación de cierto percentil que se busca. De esta manera, si quiere encontrar el trigésimo tercer percentil, utilizaría L_{33} ; y si buscara la mediana, el percentil 50o., entonces L_{50} . El número de observaciones es n ; por lo tanto, si desea localizar la mediana, su posición se encuentra en $(n + 1)/2$, o podría escribir esta expresión como $(n + 1)(P/100)$, en la que P representa el percentil que busca.

LOCALIZACIÓN DE UN PERCENTIL

$$L_p = (n + 1) \frac{P}{100} \quad (4-1)$$

Un ejemplo ayudará a explicar la fórmula anterior.

Ejemplo

En seguida aparecen las comisiones que ganó el último mes una muestra de 15 corredores de bolsa de la oficina de Salomon Smith Barney's Oakland, California. Esta compañía de inversiones tiene oficinas a lo largo de Estados Unidos.

\$2 038	\$1 758	\$1 721	\$1 637	\$2 097	\$2 047	\$2 205	\$1 787	\$2 287
1 940	2 311	2 054	2 406	1 471	1 460			

Localice la mediana, el primer y el tercer cuartiles de las comisiones ganadas.

Solución

El primer paso consiste en ordenar las comisiones ganadas de menor a mayor.

\$1 460	\$1 471	\$1 637	\$1 721	\$1 758	\$1 787	\$1 940	\$2 038
2 047	2 054	2 097	2 205	2 287	2 311	2 406	

El valor mediano es la observación que se encuentra en el centro. El valor central, o L_{50} , se localiza en $(n + 1)(50/100)$, en la que n representa el número de observaciones. En este caso es la posición número 8, determinada por $(15 + 1)(50/100)$. La octava comisión más grande es de \$2 038. Así que ésta es la mediana y la mitad de los corredores obtiene comisiones mayores que \$2 038, y la mitad gana menos de \$2 038.

Recordemos la definición de cuartil. Los cuartiles dividen a un conjunto de observaciones en cuatro partes iguales. Por consiguiente, 25% de las observaciones serán menores que el primer cuartil. Setenta y cinco por ciento de ellas serán menores que el tercer cuartil. Para localizar el primer cuartil, utilice la fórmula $(4-1)$, en la cual $n = 15$ y $P = 25$:

$$L_{25} = (n + 1) \frac{P}{100} = (15 + 1) \frac{25}{100} = 4$$

para localizar el tercer cuartil, $n = 15$ y $P = 75$:

$$L_{75} = (n + 1) \frac{P}{100} = (15 + 1) \frac{75}{100} = 12$$

Por lo tanto, los valores del primer y tercer cuartiles se localizan en las posiciones 4 y 12. El cuarto valor en la serie ordenada es \$1 721 y el decimosegundo es \$2 205. Éstos constituyen el primer y tercer cuartiles.



En el ejemplo anterior, la fórmula de localización arrojó un número entero. Es decir que al buscar el primer cuartil había 15 observaciones, así que la fórmula de localización indica que debería encontrar el cuarto valor ordenado. ¿Si hubiera 20 observaciones en la muestra, es decir $n = 20$, y quisiera localizar el primer cuartil? De acuerdo con la fórmula de localización $(4-1)$:

$$L_{25} = (n + 1) \frac{P}{100} = (20 + 1) \frac{25}{100} = 5.25$$

Localizaría el quinto valor en la serie ordenada y en seguida se desplazaría una distancia de 0.25 entre los valores quinto y sexto y señalaría a éste como el primer cuartil. Como en el caso de la mediana, el cuartil no necesita ser uno de los valores exactos del conjunto de datos.

Para explicarlo más a fondo, suponga que un conjunto de datos contiene los seis valores: 91, 75, 61, 101, 43 y 104. Trate de localizar el primer cuartil. Ordene los valores de menor a mayor: 43, 61, 75, 91, 101 y 104. El primer cuartil se localiza en

$$L_{25} = (n + 1) \frac{P}{100} = (6 + 1) \frac{25}{100} = 1.75$$

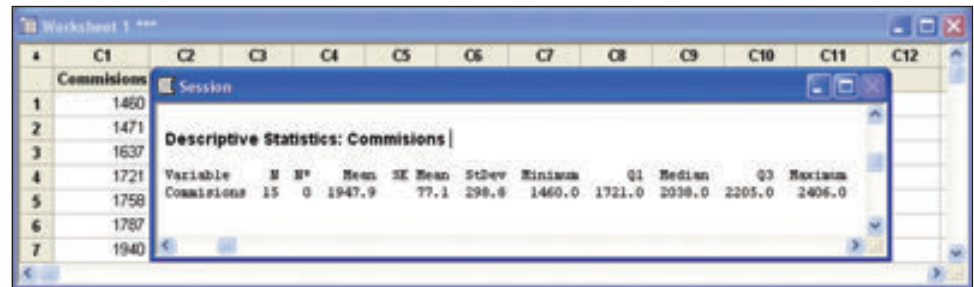
La fórmula de localización indica que el primer cuartil se ubica entre el primero y segundo valores, lo que representa 0.75 de la distancia entre ellos. El primer valor es 43 y el segundo 61. De esta manera, la distancia entre estos valores es 18. Al localizar el primer cuartil, necesita desplazarse una distancia de 0.75 entre el primero y segundo valores; así, $0.75(18) = 13.5$. Para completar el procedimiento, sume 13.5 al primer valor e indique que el primer cuartil es 56.5.

Es posible ampliar la idea para incluir tanto deciles como percentiles. Para localizar el 23o. percentil en una muestra de 80 observaciones, busque la posición 18.63.

$$L_{23} = (n + 1) \frac{P}{100} = (80 + 1) \frac{23}{100} = 18.63$$

Para determinar el valor correspondiente al 23o. percentil, localice el 18o. valor y el 19o., y determine la distancia entre ambos. Luego, multiplique esta diferencia por 0.63 y sume el resultado al valor más pequeño. El resultado sería el 23o. percentil.

Con un paquete de software de estadística, resulta relativamente sencillo ordenar los datos de menor a mayor y localizar percentiles y deciles. Tanto las salidas de Minitab como de Excel generan resúmenes estadísticos. Abajo aparece una captura de pantalla de Minitab para los datos de las comisiones de Smith Barney. Los datos incluyen el primer y el tercer cuartiles, así como la media, la mediana y la desviación. Se concluye que 25% de las comisiones fueron de menos de \$1 721 y que 75% fueron menores a \$2 205. Son los mismos valores reportados en el ejemplo previo.



The screenshot shows a Minitab Session window titled "Session". It displays the "Descriptive Statistics: Commissions" output. The background shows a worksheet with a column of commission values: 1460, 1471, 1637, 1721, 1758, 1787, 1940.

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Commissions	15	0	1947.9	77.1	298.8	1460.0	1721.0	2038.0	2205.0	2406.0

Excel y MegaStat, que se basa en Excel, calculan también los cuartiles y despliegan los resultados. Sin embargo, el método de solución que utilizan es ligeramente distinto. Para simplificar los problemas, asuma que el grupo de datos contiene un número impar de valores. El método descrito en el ejemplo, y soportado por Minitab, para el primer cuartil es:

1. Encuentre la mediana del grupo de n observaciones.
2. Concéntrese sólo en las observaciones que están *por debajo* de la mediana de estos valores. Esto es, no considere a la mediana como parte del nuevo grupo de datos.
3. Reporte este valor como el primer cuartil.

En los datos de las comisiones de Smith Barney, la comisión mediana es la octava observación en el grupo de 15 observaciones. Esta comisión es de \$2 038. La mediana de estas siete observaciones se ubica en la cuarta posición, y tiene un valor de \$1 721, el mismo valor que se encontró en el ejemplo y en la salida de Minitab.

A continuación se presenta una hoja de cálculo de Excel. También se muestran el primero y tercer cuartiles de los datos de las comisiones de Smith Barney. Note que los resultados son diferentes. Nuevamente, para simplificar la situación, asuma que existe un número impar de valores. Excel encuentra la mediana a través del siguiente método:

1. Encuentre la mediana en el grupo de n observaciones.
2. Concéntrese en todas las observaciones que son iguales a o menores que la mediana. Esto es, incluya la mediana en el nuevo subgrupo de datos.
3. Encuentre la mediana de este grupo de valores.
4. Reporte este valor como primer cuartil.

En los datos de las comisiones de Smith Barney, la mediana de las 15 observaciones originales es \$2 038. Por ello, el nuevo grupo de valores son las ocho observaciones ordenadas entre \$1 460 y \$2 038. La mediana está a medio camino entre \$1 721 y \$1 758, o \$1 739, como reportó Excel.

	A	B	C	D
1	\$1,460.00			
2	\$1,471.00			
3	\$1,637.00			
4	\$1,721.00		Cuartil 1	\$1,739.50
5	\$1,758.00			
6	\$1,787.00		Cuartil 3	\$2,151.00
7	\$1,940.00			
8	\$2,038.00			
9	\$2,047.00			
10	\$2,054.00			
11	\$2,097.00			
12	\$2,205.00			
13	\$2,287.00			
14	\$2,311.00			
15	\$2,406.00			

De manera que la diferencia esencial entre los dos métodos es:

- En el sistema Minitab, la mediana no se incluye en el subgrupo de datos.
- En el sistema Excel, la mediana se incluye en el subgrupo de datos.

En este ejemplo se consideraba un número impar de observaciones. ¿Qué pasa con el método de Excel si hay un número par de observaciones? En vez de utilizar la fórmula (4-1) para encontrar la ubicación, utiliza $0.25n + 0.75$ para descubrir la posición del primer cuartil y $0.75n + 0.25$ para hallar la posición del tercer cuartil.

¿Es importante la diferencia? No, en realidad suele ser sólo una molestia. Por lo general, los estadísticos prefieren el primer método aquí expuesto. Cuando la muestra es grande, la diferencia entre los resultados de ambos métodos es pequeña. Por ejemplo, recuerde los datos de Applewood Auto Group, que reportan la información sobre las ganancias por las ventas de 180 vehículos. A continuación se presentan los resultados de Minitab y de Excel. ¡No hay mucha diferencia, sólo \$7.00 en 180 vehículos! El reporte de cualquiera de estos valores haría muy poca diferencia en la interpretación.

Session										
Descriptive Statistics: Profit										
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Profit	180	0	1843.2	48.0	643.6	294.0	1415.5	1882.5	2275.5	3292.0

APPLEWOOD AUTO GROUP				
	A	B	C	D
1	Age	Profit		
2	44	\$294		
3	40	\$323		
4	42	\$335	Cuartil 1	1422.50
5	40	\$352	Cuartil 3	2268.50
6	46	\$369		
7	53	\$377		
8	30	\$443		
9	40	\$482		
10	37	\$732		
11	30	\$754		
12	62	\$783		
13	45	\$820		
14	50	\$842		

Autoevaluación 4-2



El departamento de control de calidad de Plainsville Peanut Company verifica el peso de un frasco de crema de cacahuete de ocho onzas. Los pesos de la muestra de nueve frascos fabricados la hora pasada son los siguientes:

7.69	7.72	7.8	7.86	7.90	7.94	7.97	8.06	8.09
------	------	-----	------	------	------	------	------	------

- ¿Cuál es el peso mediano?
- Determine los pesos correspondientes del primer y tercer cuartiles.

Ejercicios

connect™

- Determine la mediana y los valores correspondientes al primer y tercer cuartiles en los siguientes datos.

46	47	49	49	51	53	54	54	55	55	59
----	----	----	----	----	----	----	----	----	----	----

- Determine la mediana y los valores correspondientes al primer y tercer cuartiles en los siguientes datos.

5.24	6.02	6.67	7.30	7.59	7.99	8.03	8.35	8.81	9.45
9.61	10.37	10.39	11.86	12.22	12.71	13.07	13.59	13.89	15.42

- Thomas Supply Company, Inc., es un distribuidor de generadores de gas. Como en cualquier negocio, el tiempo que emplean los clientes para pagar sus recibos es importante. En la siguiente lista, en orden de menor a mayor, aparece el tiempo, en días, de una muestra de facturas de Thomas Supply Company, Inc.

13	13	13	20	26	27	31	34	34	34	35	35	36	37	38
41	41	41	45	47	47	47	50	51	53	54	56	62	67	82

- Determine el primer y tercer cuartiles.
- Determine el segundo y el octavo deciles.
- Determine el 67o. percentil.

14. Kevin Horn es el gerente nacional de ventas de National Textbooks, Inc. Cuenta con un personal de ventas conformado por 40 personas, las cuales hacen visitas a profesores universitarios en todo Estados Unidos. Cada sábado por la mañana solicita a su personal que le envíe un informe, que debe incluir, entre otras cosas, la cantidad de profesores que visitaron la semana anterior. En la lista de abajo, en orden de menor a mayor, aparece la cantidad de visitas de la semana pasada.



38	40	41	45	48	48	50	50	51	51	52	52	53	54	55	55	55	55	56	56	57
59	59	59	62	62	62	63	64	65	66	66	67	67	69	69	71	77	78	79	79	79

- Determine la cantidad mediana de visitas.
- Determine el primer y tercer cuartiles.
- Determine el primero y el noveno deciles.
- Determine el 33o. percentil.

Diagramas de caja

OA4 Construir e interpretar diagramas de caja.

Un **diagrama de caja** es una representación gráfica, basada en cuartiles, que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja, sólo necesita cinco estadísticos: el valor mínimo, Q_1 (primer cuartil), la mediana, Q_3 (tercer cuartil) y el valor máximo. Un ejemplo ayudará a explicarlo.

Ejemplo

Alexander's Pizza ofrece entregas gratuitas de pizza a 15 millas a la redonda. Alex, el propietario, desea información relacionada con el tiempo de entrega. ¿Cuánto tiempo tarda una entrega típica? ¿En qué margen de tiempo deben completarse la mayoría de las entregas? En el caso de una muestra de 20 entregas, Alex recopiló la siguiente información:

Valor mínimo = 13 minutos

Q_1 = 15 minutos

Mediana = 18 minutos

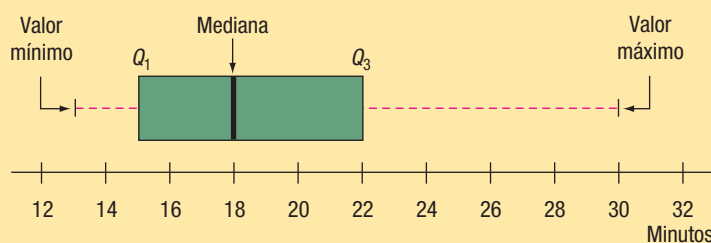
Q_3 = 22 minutos

Valor máximo = 30 minutos

Elabore un diagrama de caja de los tiempos de entrega. ¿Qué conclusiones deduce sobre los tiempos de entrega?

Solución

El primer paso para elaborar un diagrama de caja consiste en crear una escala adecuada a lo largo del eje horizontal. Luego, se debe dibujar una caja que inicie en Q_1 (15 minutos) y termine en Q_3 (22 minutos). Dentro de la caja trazamos una línea vertical para representar a la mediana (18 minutos). Por último, prolongamos líneas horizontales a partir de la caja dirigidas al valor mínimo (13 minutos) y al valor máximo (30 minutos). Estas líneas horizontales que salen de la caja, a veces reciben el nombre de *bigotes*, en virtud de que se asemejan a los bigotes de un gato.



El diagrama de caja muestra que el valor medio de las entregas, 50%, consume entre 15 y 22 minutos. La distancia entre los extremos de la caja, 7 minutos, es el **rango intercuartil**. Este rango, que es la distancia entre el primer y el tercer cuartiles, muestra la propagación o dispersión de la mayoría de las entregas.

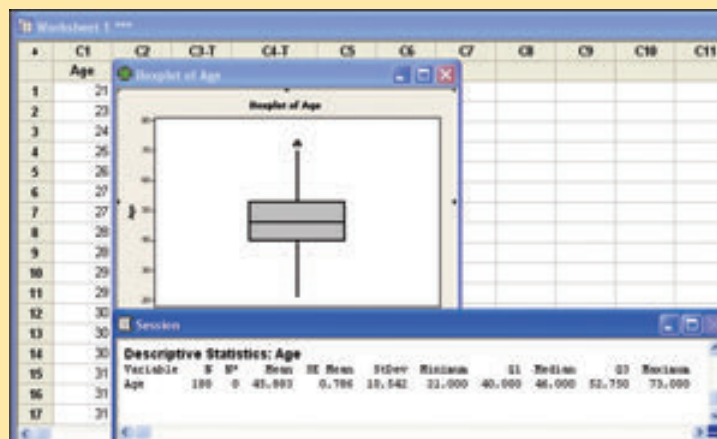
El diagrama de caja también revela que la distribución de los tiempos de entrega tiene un sesgo positivo. En el capítulo 3, página 70, recordemos que definimos el sesgo como la falta de simetría en un conjunto de datos. ¿Cómo sabe que esta distribución tiene un sesgo positivo? En este caso hay dos piezas de información que lo sugieren. Primero, la línea punteada a la derecha de la caja, que va de 22 minutos (Q_3) al tiempo máximo de 30 minutos, es más larga que la línea punteada a la izquierda que va de 15 minutos (Q_1) al valor mínimo de 13 minutos. En otras palabras, 25% de los datos mayores que el tercer cuartil se encuentran más dispersos que el 25% menor que el primer cuartil. Una segunda indicación del sesgo positivo es que la mediana no se encuentra al centro de la caja. La distancia del primer cuartil a la mediana es menor que la distancia de la mediana al tercer cuartil. El número de tiempos de entrega entre 15 y 18 minutos es el mismo que el número de tiempos de entrega entre 18 y 22 minutos.

Ejemplo

Consulte los datos de Applewood Auto Group. Elabore un diagrama de caja con base en la variable edad del comprador. ¿Cuál es la conclusión respecto de la distribución de las edades de los compradores?

Solución

Para crear el siguiente diagrama y resumen estadístico se utilizó el sistema de software de estadística de Minitab:



La edad mediana de los compradores fue de 46 años; 25% de ellos tenían menos de 40 años de edad, y 25% más de 52.75. Basándose en la información resumida y en el diagrama de caja, es posible concluir que:

- Cincuenta por ciento de los compradores están entre los 40 y los 52.75 años.
- La distribución de edades es simétrica. Existen dos razones para esta conclusión. La longitud del bigote por encima de los 52.75 años (Q_3) tiene aproximadamente el mismo largo que el bigote que está por debajo de los 40 años (Q_1). Asimismo, el área de la caja entre los 40 años y la mediana de 46 años es más o menos la misma que el área entre la mediana y los 52.75 años.

Hay tres asteriscos (*) por encima de los 70 años. ¿Qué es lo que indican? En un diagrama de caja, un asterisco identifica un **dato atípico**, es decir, que es un valor que no concuerda con el resto de los datos. Se define como un valor más de 1.5 veces la amplitud del rango intercuartil más pequeño que Q_1 , o mayor que Q_3 . En este ejemplo, un dato atípico sería un valor mayor que 71.875 años, el cual se determina con el siguiente cálculo:

$$\text{Dato atípico} > Q_3 + 1.5(Q_3 - Q_1) = 52.75 + 1.5(52.75 - 40) = 71.875$$

Un valor menor que 20.875 años también es un dato atípico.

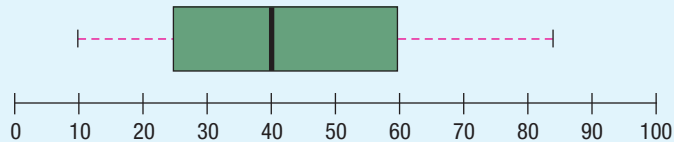
$$\text{Dato atípico} < Q_1 - 1.5(Q_3 - Q_1) = 40 - 1.5(52.75 - 40) = 20.875$$

Con base en el diagrama de caja, se concluye que hubo tres compradores de 72 años o mayores, y ninguno menor de 21 años. Nota técnica: en algunos casos, un solo asterisco puede representar más de una observación, en razón de las limitaciones del software y del espacio disponible. Es buena idea verificar los datos reales. En este caso, hubo tres compradores de 72 años o mayores: dos tienen 72 y uno tiene 73.

Autoevaluación 4-3



El siguiente diagrama de caja muestra los activos en millones de dólares de cooperativas de crédito en Seattle, Washington.

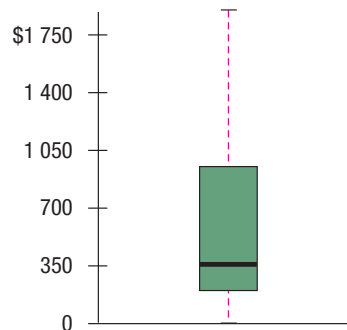


¿Cuáles son los valores mínimo y máximo, los cuartiles primero y tercero, y la mediana? ¿Estaría usted de acuerdo en que la distribución es simétrica? ¿Hay datos atípicos?

Ejercicios

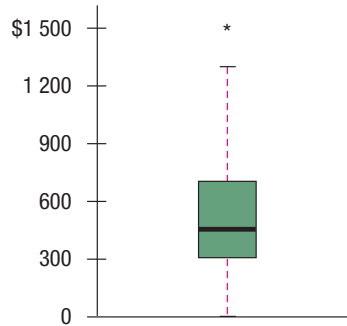
connect™

15. El diagrama de caja muestra la suma que se gastaron en libros y suministros durante un año los estudiantes de cuarto año de universidades públicas.



- Calcule la mediana de la suma que se gastó.
- Calcule el primero y el tercer cuartiles de la cantidad que se gastó.
- Calcule el rango intercuartil de la cantidad que se gastó.
- ¿Más allá de qué punto un valor se considera dato atípico?

- e) Identifique cualesquiera datos atípicos y calcule su valor.
 f) ¿Es la distribución simétrica, o tiene sesgo positivo o negativo?
16. El diagrama de caja muestra el cargo interestatal de crédito por hora para carreras de cuatro años de estudiantes graduados en universidades públicas.



- a) Calcule la mediana.
 b) Calcule el primer y tercer cuartiles.
 c) Determine el rango intercuartil.
 d) ¿Más allá de qué punto se considera dato atípico un valor?
 e) Identifique cualesquiera datos atípicos y calcule su valor.
 f) ¿La distribución es simétrica, o tiene sesgo positivo o negativo?
17. En un estudio sobre el rendimiento en millas por galón de gasolina de automóviles modelo 2011, la media fue de 27.5 y la mediana de 26.8. El valor más pequeño fue de 12.70 millas por galón y el más grande de 50.20. El primer y tercer intercuartiles fueron 17.95 y 35.45 millas por galón, respectivamente. Elabore un diagrama de caja y haga algún comentario sobre la distribución. ¿Es una distribución simétrica?
18. Una muestra de 28 departamentos de tiempo compartido en el área de Orlando, Florida, reveló las siguientes tarifas diarias de una suite con una recámara. Por comodidad, los datos se encuentran ordenados de menor a mayor. Construya un diagrama de caja para representar los datos. Haga algún comentario sobre la distribución. Identifique el primer y tercer cuartiles, y la mediana. 

\$116	\$121	\$157	\$192	\$207	\$209	\$209
229	232	236	236	239	243	246
260	264	276	281	283	289	296
307	309	312	317	324	341	353

4.5 Sesgo

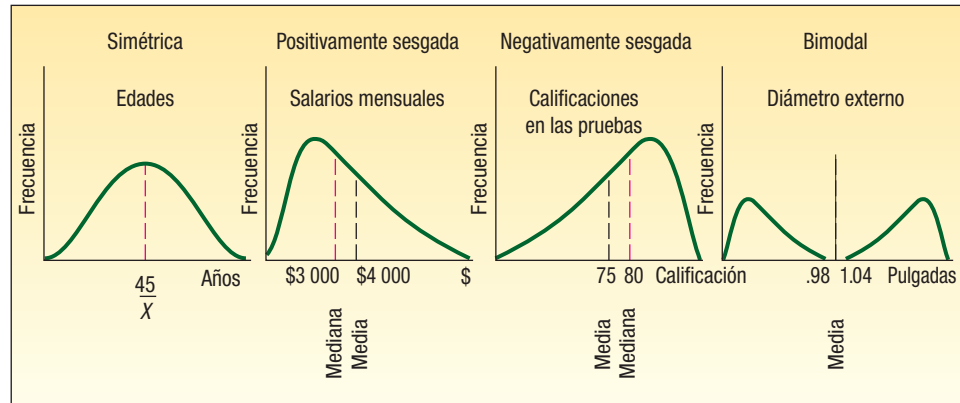
En el capítulo 3 se trataron las medidas de ubicación central de un conjunto de observaciones por medio de la presentación de un informe sobre la media, la mediana y la moda. También se describieron medidas que muestran el grado de propagación o variación de un conjunto de datos, como el rango y la desviación estándar.

Otra característica de un conjunto de datos es la forma. Hay cuatro formas: simétrica, con sesgo positivo, con sesgo negativo y bimodal. En un conjunto **simétrico** de observaciones la media y la mediana son iguales, y los valores de datos se dispersan uniformemente en torno a estos valores. Los valores de datos debajo de la media y de la mediana constituyen una imagen especular de los datos arriba de estas medidas. Un conjunto de valores se encuentra **sesgado a la derecha o positivamente sesgado** si existe un solo pico y los valores se extienden mucho más allá a la derecha del pico que a la izquierda de éste. En este caso la media es más grande que la mediana. En una distribución **negativamente sesgada** existe un solo pico, pero las observaciones se extienden más a la izquierda, en dirección negativa. En una distribución negativamente sesgada, la media es menor que la mediana. Las distribuciones positivamente

OA5 Calcular y entender el coeficiente de sesgo.

El sesgo muestra la falta de simetría en un grupo de observaciones.

sesgadas son más comunes. Con frecuencia, los salarios obedecen este patrón. Piense en los salarios del personal de una pequeña compañía con alrededor de 100 empleados. El presidente y unos cuantos altos ejecutivos recibirían mucho más que los demás trabajadores, por lo que la distribución de salarios mostraría un sesgo positivo. Una **distribución bimodal** tendrá dos o más picos. Con frecuencia éste es el caso cuando los valores provienen de dos o más poblaciones. Esta información se resume en la gráfica 4-1.



GRÁFICA 4-1 Formas de los polígonos de frecuencias



Estadística en acción

El difunto Stephen Jay Gould (1941-2002) fue profesor de zoología y de geología en la Universidad de Harvard. En 1982 se le diagnosticó cáncer y le dieron ocho meses de vida. No obstante, y sin darse por vencido, mostró en su investigación que la distribución de tiempos de supervivencia se encuentra drásticamente sesgada a la derecha y que no sólo 50% de pacientes de cáncer similar sobreviven más de 8 meses, sino que el tiempo de supervivencia podía ser de años, no de meses. Sobre la base de su experiencia, escribió un ensayo varias veces publicado que se tituló "The Median Is not the Message" (La mediana no es el mensaje).

En la literatura estadística se utilizan diversas fórmulas para calcular el sesgo. La más sencilla, ideada por el profesor Karl Pearson (1857-1936), se basa en la diferencia entre la media y la mediana.

COEFICIENTE DE SESGO DE PEARSON

$$sk = \frac{3(\bar{X} - \text{Mediana})}{s} \quad (4-2)$$

De acuerdo con esta expresión, el sesgo puede variar de -3 a 3 . Un valor próximo a -3 , como -2.57 , indica un sesgo negativo considerable. Un valor como 1.63 indica un sesgo positivo moderado. Un valor de 0 , que ocurre cuando la media y la mediana son iguales, indica que la distribución es simétrica y que no se presenta ningún sesgo.

En esta obra se presentan resultados que se obtuvieron con paquetes de software de estadística en Minitab y Excel. Con ambos se calcula un valor del coeficiente de sesgo basado en las desviaciones de la media elevadas al cubo. La fórmula es la siguiente:

COEFICIENTE DE SESGO CALCULADO CON SOFTWARE

$$sk = \frac{n}{(n-1)(n-2)} \left[\sum \left(\frac{X - \bar{X}}{s} \right)^3 \right] \quad (4-3)$$

La fórmula (4-3) permite comprender la idea de sesgo. El miembro derecho de la fórmula es la diferencia entre cada valor y la media, dividida entre la desviación estándar. Esto corresponde a la porción $(X - \bar{X})/s$ de la fórmula. Esta idea recibe el nombre de **estandarización**. El concepto de estandarización de un valor se analiza con más detalle en el capítulo 7 cuando se describe la distribución de probabilidad normal. En este punto, observe que el resultado

consiste en la diferencia entre cada valor y la media en unidades de desviación estándar. Si la diferencia es positiva, el valor particular es más grande que la media; si la variación es negativa, la cantidad estandarizada es menor que la media. Cuando eleva al cubo estos valores, conserva la información relativa a la diferencia. Recuerde que en la fórmula de la desviación estándar [vea fórmula (3-11)], se elevó al cuadrado la diferencia entre cada valor y la media de tal manera que, como resultado, todos los valores eran no negativos.

Si el conjunto de valores de datos que se estudia es simétrico, al elevar al cubo los valores estandarizados y sumar todos los valores, el resultado se aproximaría a cero. Si hay varios valores grandes, claramente separados unos de otros, la suma de las diferencias al cubo sería un valor positivo grande. Valores mucho menores dan como resultado una suma al cubo negativa.

Un ejemplo ilustrará la idea de sesgo.

Ejemplo

En seguida aparecen las utilidades por acción que obtuvo una muestra de 15 compañías de software durante el año 2010. Las utilidades por acción se encuentran ordenadas de menor a mayor.

\$0.09	\$0.13	\$0.41	\$0.51	\$ 1.12	\$ 1.20	\$ 1.49	\$3.18
3.50	6.36	7.83	8.92	10.13	12.99	16.40	

Calcule la media, la mediana y la desviación estándar. Determine el coeficiente de sesgo utilizando los métodos de Pearson y de software. ¿Qué concluye respecto de la forma de la distribución?

Solución

Éstos son los datos de la muestra, así que aplique la fórmula (3-2) para determinar la media:

$$\bar{X} = \frac{\sum X}{n} = \frac{\$74.26}{15} = \$4.95$$

La mediana es el valor intermedio de un conjunto de datos, ordenados de menor a mayor. En este caso, el valor medio es \$3.18, así la mediana de las utilidades por acción es \$3.18.

Emplee la fórmula (3-11) de la página 84 para calcular la desviación estándar de la muestra:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{(\$0.09 - \$4.95)^2 + \cdots + (\$16.40 - \$4.95)^2}{15 - 1}} = \$5.22$$

El coeficiente de sesgo de Pearson es de 1.017, calculado de la siguiente manera:

$$sk = \frac{3(\bar{X} - \text{Mediana})}{s} = \frac{3(\$4.95 - \$3.18)}{\$5.22} = 1.017$$

Esto indica que existe un sesgo positivo moderado en los datos de las utilidades por acción.

Cuando se utiliza el método del software resulta un valor similar, aunque no exactamente el mismo. Los detalles de los cálculos aparecen en la tabla 4-2. Para comenzar, determine la diferencia entre las utilidades por acción, así como la media, y divida el resultado entre la desviación estándar. Recuerde que a esto se llama *estandarización*. Luego, eleve al cubo, es decir, eleve a la tercera potencia el resultado del primer paso. Por último, sume los valores elevados al cubo. Los detalles en el caso de la primera compañía, es decir, en la compañía con utilidades de \$0.09 por acción, son:

$$\left(\frac{X - \bar{X}}{s}\right)^3 = \left(\frac{0.09 - 4.95}{5.22}\right)^3 = (-0.9310)^3 = -0.8070$$

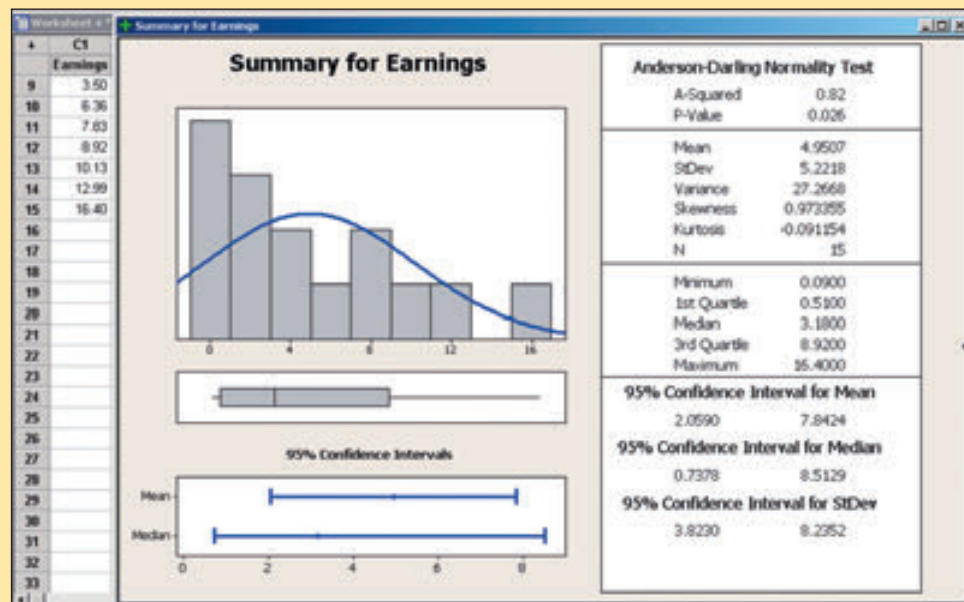
TABLA 4-2 Cálculo del coeficiente de sesgo

Utilidades por acción	$\frac{(X - \bar{X})}{s}$	$\left(\frac{X - \bar{X}}{s}\right)^3$
0.09	-0.9310	-0.8070
0.13	-0.9234	-0.7873
0.41	-0.8697	-0.6579
0.51	-0.8506	-0.6154
1.12	-0.7337	-0.3950
1.20	-0.7184	-0.3708
1.49	-0.6628	-0.2912
3.18	-0.3391	-0.0390
3.50	-0.2778	-0.0214
6.36	0.2701	0.0197
7.83	0.5517	0.1679
8.92	0.7605	0.4399
10.13	0.9923	0.9772
12.99	1.5402	3.6539
16.40	2.1935	10.5537
		11.8274

Cuando sume los 15 valores cúbicos, el resultado es 11.8274. Es decir, el término $\sum[(X - \bar{X})/s]^3 = 11.8274$. Para determinar el coeficiente de sesgo, utilice la fórmula (4-3), con $n = 15$.

$$sk = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X - \bar{X}}{s} \right)^3 = \frac{15}{(15-1)(15-2)} (11.8274) = 0.975$$

La conclusión es que los valores de las utilidades por acción se encuentran un tanto sesgadas positivamente. El siguiente diagrama, de Minitab, muestra las medidas descriptivas, como la media, la mediana y la desviación estándar de los datos por utilidades por acción. Incluye, asimismo, el coeficiente de sesgo y un histograma con una curva con forma de campana superpuesta.



Autoevaluación 4-4



Una muestra de cinco capturistas de datos que laboran en la oficina de impuestos de Horry County revisó el siguiente número de expedientes fiscales durante la última hora: 73, 98, 60, 92 y 84.

- Calcule la media, la mediana y la desviación estándar.
- Calcule el coeficiente de sesgo con el método de Pearson.
- Calcule el coeficiente de sesgo usando un paquete de software.
- ¿Qué conclusión obtiene respecto del sesgo de los datos?

Ejercicios

connect™

En el caso de los ejercicios 19-22:

- Calcule la media, la mediana y la desviación estándar.
- Calcule el coeficiente de sesgo con el método de Pearson.
- Estime el coeficiente de sesgo con un paquete de software.

19. Los siguientes valores son los sueldos iniciales, en miles de dólares, de una muestra de cinco graduados de contabilidad, quienes aceptaron puestos de contaduría pública el año pasado. 

36.0	26.0	33.0	28.0	31.0
------	------	------	------	------

20. En la siguiente lista aparecen los salarios, en miles de dólares, de una muestra de 15 directores de finanzas de la industria electrónica. 

\$516.0	\$548.0	\$566.0	\$534.0	\$586.0	\$529.0
546.0	523.0	538.0	523.0	551.0	552.0
486.0	558.0	574.0			

21. A continuación aparece una lista de las comisiones (en miles de dólares) que percibieron el año pasado los representantes de ventas de Furniture Patch, Inc. 

\$ 3.9	\$ 5.7	\$ 7.3	\$10.6	\$13.0	\$13.6	\$15.1	\$15.8	\$17.1
17.4	17.6	22.3	38.6	43.2	87.7			

22. La lista que sigue está conformada por los salarios de los 25 jugadores en la nómina del día de la apertura de los Yankees de Nueva York en 2010. La información de los salarios se expresa en miles de dólares. 

Jugador	Salario (miles de dólares)	Posición	Jugador	Salario (miles de dólares)	Posición
Aceves, Alfredo	435.7	Pitcher	Park, Chan Ho	1 200.0	Pitcher
Burnett, A.J.	16 500.0	Pitcher	Pena, Ramiro	412.1	Defensa
Cano, Robinson	9 000.0	Segunda base	Pettitte, Andy	11 750.0	Pitcher
Cervelli, Francisco	410.8	Catcher	Posada, Jorge	13 100.0	Catcher
Chamberlain, Joba	488.0	Pitcher	Rivera, Mariano	15 000.0	Pitcher
Gardner, Brett	452.5	Jardinero	Robertson, David	426.7	Pitcher
Granderson, Curtis	5 500.0	Jardinero	Rodriguez, Alex	33 000.0	Tercera base
Hughes, Phil	447.0	Pitcher	Sabathia, CC	24 285.7	Pitcher
Jeter, Derek	22 600.0	Receptor	Swisher, Nick	6 850.0	Jardinero
		de pase corto	Teixeira, Mark	20 625.0	Primera base
Johnson, Nick	5 500.0	Primera base	Thames, Marcus	900.0	Jardinero
Marte, Damaso	4 000.0	Pitcher	Vazquez, Javier	11 500.0	Pitcher
Mitre, Sergio	850.0	Pitcher	Winn, Randy	1 100.0	Jardinero

4.6 Descripción de la relación entre dos variables



En el capítulo 2 y en la primera sección de éste se han expuesto técnicas gráficas para resumir la distribución de una sola variable. En el capítulo 2 se empleó un histograma para resumir las ganancias por vehículos vendidos en Applewood Auto Group. En este capítulo las herramientas que se usaron fueron los diagramas de puntos y las gráficas de tallo y hojas para representar visualmente un conjunto de datos. En tanto que aparece una sola variable, se habla de datos **univariados**.

Hay situaciones en las que se estudia y representa visualmente la relación entre dos variables. Al estudiar la relación entre ellas, se hace referencia a los datos como **bivariados**. Con frecuencia, los analistas de datos tratan de entender la relación entre dos variables. He aquí algunos ejemplos:

- Tybo and Associates es una firma de abogados que se anuncia mucho en televisión. Los socios están considerando la forma de incrementar su presupuesto publicitario. Antes de hacerlo, les gustaría conocer la relación entre la cantidad que se gasta al mes en publicidad y la cantidad total de cuentas por cobrar en dicho mes. En otras palabras, ¿un incremento de la suma que se gasta en publicidad dará como resultado un incremento de las cuentas por cobrar?
- Coastal Realty estudia sus precios de venta de casas. ¿Qué variables parecen estar relacionadas con ellos? Por ejemplo, ¿las casas más grandes se venden a un precio superior que las más pequeñas? Es probable. Por ello, Coastal tendría que estudiar la relación entre el área en pies cuadrados y el precio de venta.
- El doctor Stephen Givens es experto en desarrollo humano. Estudia la relación entre la altura de los padres y la de sus hijos. Es decir, ¿los padres altos tienden a tener hijos altos? ¿Esperaría usted que Shaquille O'Neal, el jugador de baloncesto profesional de siete pies y una pulgada de altura y 335 libras de peso tuviera hijos relativamente altos?

Una técnica gráfica útil para mostrar la relación entre variables es el **diagrama de dispersión**.

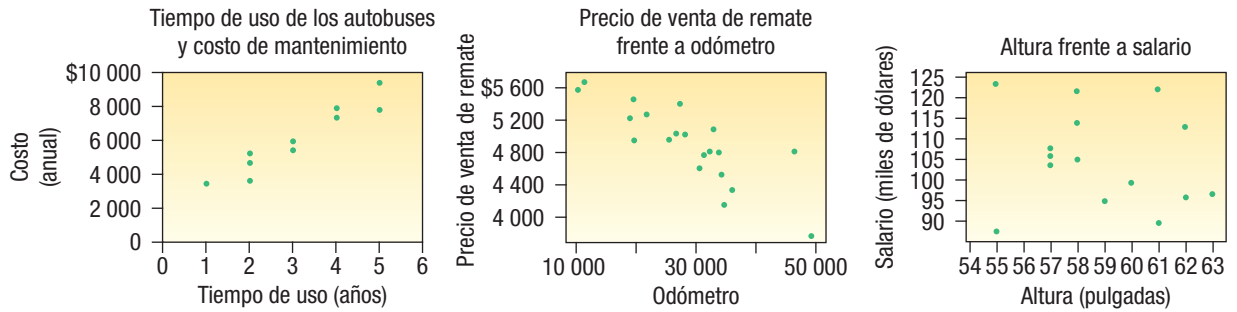
OA6 Trazar e interpretar un diagrama de dispersión.

Para trazar un diagrama de dispersión son necesarias dos variables. Se escala una de las variables sobre el eje horizontal (eje X) de una gráfica y la otra variable a lo largo del eje vertical (eje Y). Por lo general, una de las variables depende hasta cierto grado de la otra. En el tercer ejemplo citado, la altura del hijo *depende* de la altura del padre. Así que se representa la altura del padre en el eje horizontal y la del hijo sobre el eje vertical.

Un software de estadística, como Excel, sirve para ejecutar la función de trazo. *Precaución:* siempre se debe tener cuidado en la escala. Al cambiar la escala, ya sea del eje vertical o del eje horizontal, se afecta la fuerza de la relación visual.

El diagrama de dispersión se usa como forma de entender la relación entre dos variables.

A continuación aparecen tres diagramas de dispersión (gráfica 4-2). El de la izquierda muestra una mayor relación entre el tiempo de uso y el costo de mantenimiento durante el año pasado de una muestra de 10 autobuses propiedad de la ciudad de Cleveland, Ohio. Observe que a medida que se incrementa el tiempo de uso del autobús, también aumenta el costo anual de mantenimiento. El ejemplo del centro, relativo a una muestra de 20 vehículos, muestra una fuerte relación indirecta entre la lectura del odómetro y el precio de venta de remate. Es decir, conforme aumente el número de millas recorridas, el precio de venta de remate se reduce. El ejemplo de la derecha describe la relación entre la altura y el salario anual de una muestra de 15 supervisores de turno. Esta gráfica indica que existe poca relación entre la altura y el salario anual.



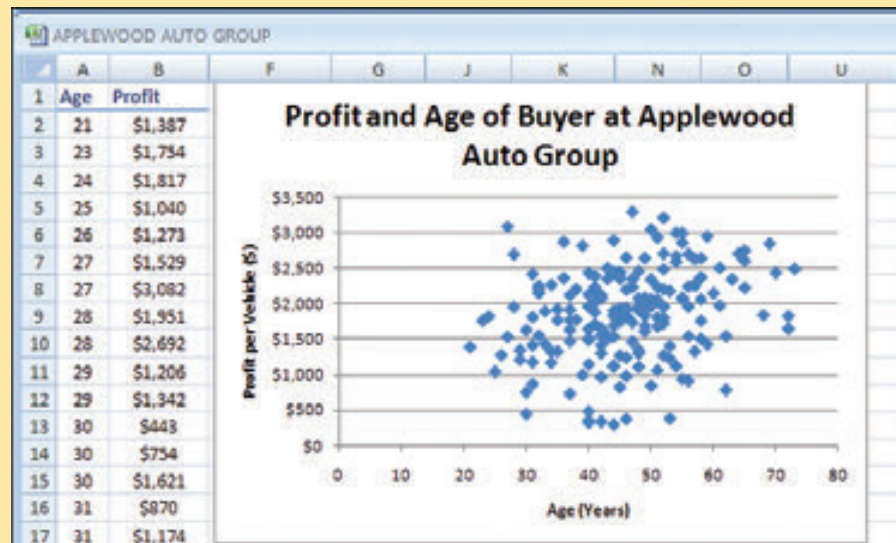
GRÁFICA 4-2 Tres ejemplos de diagramas de dispersión

Ejemplo

En la introducción del capítulo 2 aparecen datos de Applewood Auto Group. Se reunió información sobre diversas variables, entre ellas la ganancia que se obtuvo por la venta de 180 vehículos el mes pasado. Además del monto de la ganancia en cada venta, otra de las variables es la edad del comprador. ¿Existe alguna relación entre la ganancia que se obtuvo por la venta de un vehículo y la edad del comprador? ¿Sería razonable concluir que se gana más en los vehículos que adquieren los compradores de más edad?

Solución

Es posible investigar la relación entre la ganancia por vehículo vendido y la edad del comprador con un diagrama de dispersión. Represente la escala de edad sobre el eje horizontal, o eje X, y la ganancia sobre el eje vertical, o eje Y. Utilice Microsoft Excel para crear un diagrama de dispersión. Los comandos de Excel necesarios para la captura de pantalla se muestran en la sección **Comandos de software** ubicada al final del capítulo.



El diagrama de dispersión muestra una relación positiva entre las dos variables. No parece haber mucha relación entre la ganancia por vehículo y la edad del comprador. En el capítulo 13 estudiaremos más ampliamente la relación entre variables, incluso calcularemos varias medidas numéricas para expresar la relación entre variables.

En el ejemplo anterior hay una débil relación positiva, o directa, entre las variables. Sin embargo, hay muchos casos en los que existe una relación entre las variables, pero dicha relación es inversa o negativa. Por ejemplo:

- El valor de un vehículo y el número de millas recorridas. Conforme la cantidad de millas se incrementa, el valor del vehículo desciende.
- La prima de un seguro de automóvil y la edad del conductor. Las cuotas de automóvil tienden ser las más altas para los adultos jóvenes y menores para personas de más edad.
- En el caso de muchos oficiales encargados de hacer que se cumpla la ley, conforme aumenta el número de años de trabajo, la cantidad de multas de tránsito disminuye. Esto puede deberse a que el personal se torna más liberal en sus interpretaciones o a que quizá tengan puestos de supervisión y no un cargo en el que puedan levantar tantas multas. Pero en cualquier caso, conforme la edad aumenta, la cantidad de multas se reduce.

Un diagrama de dispersión requiere que las dos variables sean por lo menos de escala de intervalo. En el ejemplo de Applewood Auto Group, tanto la edad como la ganancia de la venta son variables de escala de razón. La altura también es una escala de razón, según la manera en la que se utilizó en el estudio de la relación entre la altura de los padres y la de los hijos. ¿Y si desea estudiar la relación entre dos variables cuando una o ambas son de escala nominal u ordinal? En este caso, debe registrar los resultados en una **tabla de contingencia**.

OA7 Construir e interpretar una tabla de contingencia.

TABLA DE CONTINGENCIA Tabla que se utiliza para clasificar observaciones de acuerdo con dos características identificables.

Una tabla de contingencia es una tabulación cruzada, que resume simultáneamente dos variables de interés. Por ejemplo:

- Los estudiantes en una universidad se clasifican por género y lugar en la clase.
- Un producto se clasifica como aceptable o inaceptable y de acuerdo con el turno (matutino, vespertino, nocturno) en el que se le fabrica.
- Un votante de una escuela que lleva a cabo un referendo para otorgar becas se clasifica de acuerdo con su afiliación partidista (demócrata, republicano u otro), y el número de hijos que asisten a la escuela del distrito (0, 1, 2, etcétera).

Ejemplo

Solución

Hay cuatro distribuidoras en el Applewood Auto Group. Suponga que desea comparar la ganancia que se obtuvo por cada vehículo vendido por una concesionaria en particular. Dicho de otra forma, ¿existe una relación entre el monto de ganancia y la distribuidora?

El nivel de medida de la variable concesionaria es nominal y de razón en el caso de la variable ganancia. Para usar con eficiencia una tabla de contingencias, ambas variables deben ser ya sea nominales u ordinales. Para hacer que las variables sean compatibles, hay que clasificar la variable ganancia en dos categorías: aquellos casos en los que la ganancia que se obtuvo es mayor a la mediana, y aquellos en que es menor. En la página 69 se calculó que la ganancia mediana por todas las ventas del mes pasado en Applewood Auto Group es de \$1 882.50.

Tabla de contingencia sobre la relación entre ganancia y concesionaria					
Abajo/arriba					
Ganancia mediana	Kane	Olean	Sheffield	Tionesta	Total
Por abajo	25	20	19	26	90
Por arriba	27	20	26	17	90
Total	52	40	45	43	180

Si se organiza la información en una tabla de contingencia, es posible comparar la ganancia de las cuatro distribuidoras. Se observa lo siguiente:

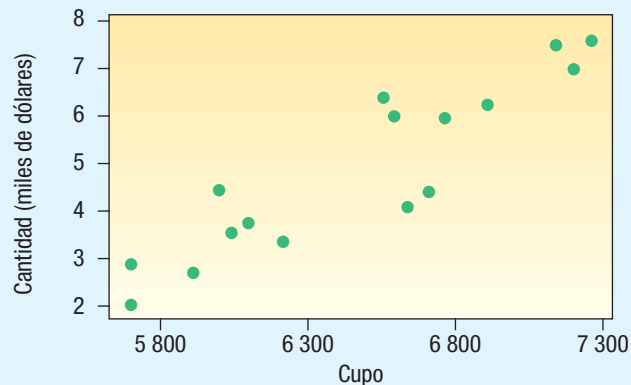
- De la columna Total a la derecha, 90 de los 180 autos vendidos dieron una ganancia por encima de la mediana, y la otra mitad, por debajo. Esto era lo esperado, dada la definición de mediana.
- En el caso de la distribuidora Kane, 25 de los 52 vehículos, o 48%, fueron vendidos con una ganancia mayor a la mediana.
- El porcentaje de ganancias por encima de la mediana de las otras concesionarias es 50% en el caso de Olean, 42% en el de Sheffield y 60% en el de Tionesta.

Volveremos al estudio de las tablas de contingencia en el capítulo 5 cuando veamos la probabilidad, y en el capítulo 17 cuando estudiemos los métodos no paramétricos de análisis.

Autoevaluación 4-5



El grupo de rock Blue String Beans está de gira por Estados Unidos. El siguiente diagrama muestra la relación entre el cupo para el concierto y el ingreso en miles de dólares en una muestra de conciertos.



- ¿Qué nombre recibe el diagrama?
- ¿Cuántos conciertos se estudiaron?
- Calcule los ingresos de un concierto con lleno total.
- ¿Cómo caracterizaría la relación entre ingresos y cupo? ¿Es fuerte o débil, directa o inversa?

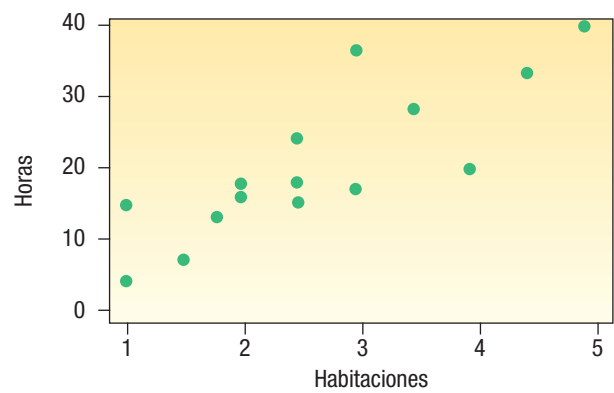
Ejercicios

connect™

23. Elabore el diagrama de dispersión de los siguientes datos tomados de una muestra. ¿Cómo describiría la relación entre los valores? 

Valor X	Valor Y	Valor X	Valor Y
10	6	11	6
8	2	10	5
9	6	7	2
11	5	7	3
13	7	11	7

24. Silver Springs Moving and Storage, Inc., estudia la relación que existe entre el número de habitaciones en una mudanza y el número de horas que se requieren de trabajo para completarla. Como parte del análisis, el director de finanzas de Silver Springs creó el siguiente diagrama de dispersión.



- a) ¿Cuántas mudanzas se incluyen en la muestra?
b) ¿Parece que se requieren más horas de trabajo si la cantidad de habitaciones se incrementa, o las horas de trabajo disminuyen si aumenta la cantidad de habitaciones?
25. El director de planeación de Devine Dining, Inc., desea estudiar la relación entre el género de un huésped y si éste ordena postre. Para investigar esta relación, recopiló la siguiente información de 200 consumidores.

Orden de postre	Género		Total
	Hombre	Mujer	
Sí	32	15	47
No	68	85	153
Total	100	100	200

- a) ¿Cuál es el nivel de medición de las dos variables?
b) ¿Qué nombre recibe esta tabla?
c) A partir de la evidencia que ofrece la tabla, ¿los hombres piden más postre que las mujeres? Explique su respuesta.
26. Sky Resorts Inc., de Vermont, considera su fusión con Gulf Shores, Inc., de Alabama. El consejo directivo encuestó a 50 accionistas acerca de su posición sobre la fusión. Los resultados aparecen en seguida.

Número de participación	Opinión			Total
	A favor	En contra	Indeciso	
Menos de 200	8	6	2	16
200 a 1 000	6	8	1	15
Más de 1 000	6	12	1	19
Total	20	26	4	50

- a) ¿Cuál es el nivel de medición que se empleó en la tabla?
b) ¿Qué nombre recibe esta tabla?
c) ¿Qué grupo parece oponerse con más fuerza a la fusión?

Resumen del capítulo

- I. Un diagrama de puntos muestra el rango de valores sobre el eje horizontal, y se coloca un punto por encima de cada uno de los valores.
 - A. Un diagrama de puntos muestra los detalles de cada observación.
 - B. Es de utilidad para comparar dos o más conjuntos de datos.
- II. Un diagrama de tallo y hojas constituye una alternativa al histograma.
 - A. El dígito principal es el tallo y el dígito secundario, la hoja.
 - B. Las ventajas de un diagrama de tallo y hojas sobre un histograma incluyen las siguientes:
 1. La identidad de cada observación no se pierde.
 2. Los dígitos proporcionan una representación de la distribución.
 3. También se exhiben las frecuencias acumulativas.
- III. Las medidas de localización describen la forma de un conjunto de observaciones.
 - A. Los cuartiles dividen un conjunto de observaciones en cuatro partes iguales.
 1. Veinticinco por ciento de las observaciones son menores que el primer cuartil, 50% son menores que el segundo cuartil y 75% son menores que el tercer cuartil.
 2. El rango intercuartil es la diferencia entre el tercer y el primer cuartiles.
 - B. Los deciles dividen a un conjunto de observaciones en diez partes iguales y los percentiles en 100 partes iguales.
 - C. Un diagrama de caja es una representación gráfica de un conjunto de datos.
 1. Se traza una caja que encierra las regiones entre el primer y tercer cuartiles.
 - a) Se dibuja una línea en el interior de la caja en el valor intermedio.
 - b) Los segmentos punteados se prolongan a partir del tercer cuartil hasta el valor más alto con el fin de mostrar el 25% más alto y a partir del primer cuartil hasta el valor más bajo con el fin de mostrar el 25% más bajo de los valores.
 2. Un diagrama de caja se basa en cinco estadísticos: los valores máximo y mínimo, el primer y tercer cuartiles y la mediana.
- IV. El coeficiente de sesgo es una medida de la simetría de una distribución.
 - A. Existen dos fórmulas para determinar el coeficiente de sesgo.
 1. La fórmula que elaboró Pearson es:

$$sk = \frac{3(\bar{X} - \text{Mediana})}{s} \quad (4-2)$$

2. El coeficiente de sesgo calculado con un software de estadística es:

$$sk = \frac{n}{(n-1)(n-2)} \left[\sum \left(\frac{X - \bar{X}}{s} \right)^3 \right] \quad (4-3)$$

- V. Un diagrama de dispersión es una herramienta gráfica para representar la relación entre dos variables.
 - A. Ambas variables se miden con escalas de intervalo o de razón.
 - B. Si la propagación de los puntos se dirige de la parte inferior izquierda a la parte superior derecha, las variables que se estudian se encuentran directa o positivamente relacionadas.
 - C. Si la dispersión de los puntos se orienta de la parte superior izquierda a la inferior derecha, las variables se encuentran relacionadas inversa o negativamente.
- VI. Una tabla de contingencia se utiliza para clasificar observaciones de escala nominal de acuerdo con dos características.

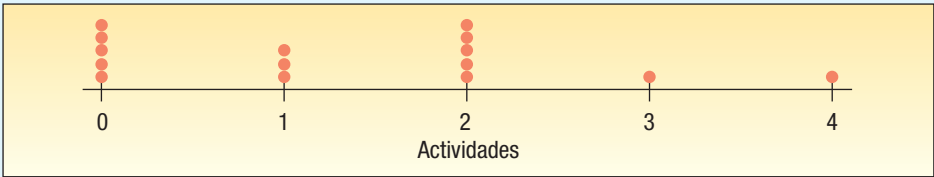
Clave de pronunciación

SÍMBOLO	SIGNIFICADO	PRONUNCIACIÓN
L_p	Ubicación del percentil	<i>L subíndice p</i>
Q_1	Primer cuartil	<i>Q subíndice 1</i>
Q_3	Tercer cuartil	<i>Q subíndice 3</i>

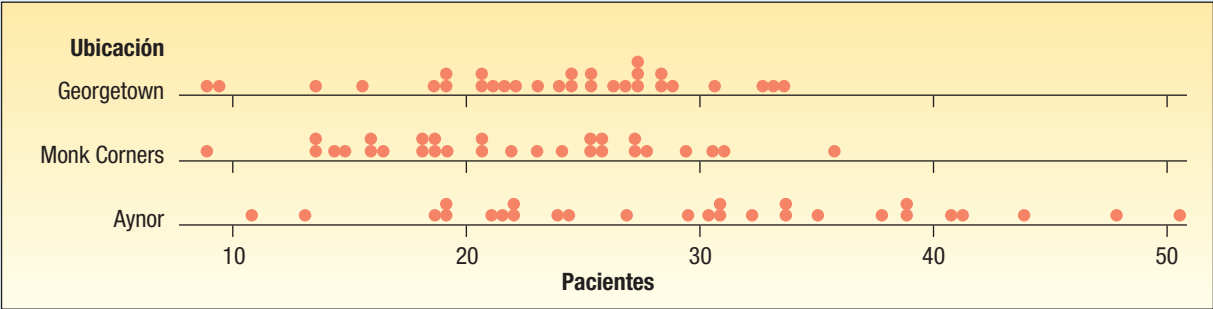


Ejercicios del capítulo

27. Se le preguntó a una muestra de estudiantes que asiste a la Southeast Florida University por la cantidad de actividades sociales en las que participaron la semana pasada. El diagrama que aparece en seguida se construyó a partir de datos tomados de la muestra.



- a) ¿Cuál es el nombre que se da a este diagrama?
 - b) ¿Cuántos estudiantes se incluyeron en el estudio?
 - c) ¿Cuántos estudiantes informaron que no asistían a ninguna actividad social?
28. Doctor's Care es una clínica ambulatoria que tiene sucursales en Georgetown, Monks Corners y Aynor, y en la cual los pacientes reciben tratamiento por lesiones menores, resfriados, gripes y se les practican exámenes físicos. Los siguientes diagramas muestran la cantidad de pacientes que se trataron en las tres sucursales el mes pasado.



Describa el número de pacientes atendidos en las tres sucursales cada día. ¿Cuáles son los números máximo y mínimo de pacientes que se atendieron en cada una de las sucursales?

29. A continuación se proporciona el tamaño de la pantalla de 23 televisores LCD. Elabore un diagrama de tallos y hojas de esta variable

46	52	46	40	42	46	40	37	46	40	52	32	37	32	52
40	32	52	40	52	46	46	52							

30. La siguiente tabla muestra las 25 compañías (ordenadas por capitalización del mercado) que operan en el área de Washington, DC, junto al año en que fueron fundadas y el número de empleados. Elabore un diagrama de tallo y hojas de estas variables y escriba una breve descripción de sus hallazgos.

Compañía	Año de fundación	Empleados
AES Corp.	1981	30 000
American Capital Strategies Ltd.	1986	484
AvalonBay Communities Inc.	1978	1 767
Capital One Financial Corp.	1995	31 800
Constellation Energy Group Inc.	1816	9 736
Coventry Health Care Inc.	1986	10 250
Danaher Corp.	1984	45 000
Dominion Resources Inc.	1909	17 500
Fannie Mae	1938	6 450
Freddie Mac	1970	5 533
		(continúa)

Compañía	Año de fundación	Empleados
Gannett Co.	1906	49 675
General Dynamics Corp.	1952	81 000
Genworth Financial Inc.	2004	7 200
Harman International Industries Inc.	1980	11 246
Host Hotels & Resorts Inc.	1927	229
Legg Mason Inc.	1899	3 800
Lockheed Martin Corp.	1995	140 000
Marriott International Inc.	1927	151 000
MedImmune Inc.	1988	2 516
NII Holdings Inc.	1996	7 748
Norfolk Southern Corp.	1982	30 594
Pepco Holdings Inc.	1896	5 057
Sallie Mae	1972	11 456
Sprint Nextel Corp.	1899	64 000
T. Rowe Price Group Inc.	1937	4 605
The Washington Post Co.	1877	17 100

31. En años recientes, como consecuencia de las bajas tasas de interés, muchos propietarios de casas refinanciaron sus créditos. Linda Lahey es agente hipotecaria de Down River Federal Savings and Loan. A continuación aparecen las sumas refinanciadas de 20 préstamos a los que les dio curso la semana pasada. Los datos se expresan en miles de dólares y se encuentran ordenados de menor a mayor. 

59.2	59.5	61.6	65.5	66.6	72.9	74.8	77.3	79.2
83.7	85.6	85.8	86.6	87.0	87.1	90.2	93.3	98.6
100.2	100.7							

- a) Calcule la mediana, el primer cuartil y el tercer cuartil.
b) Determine los percentiles 26o. y 83o.
c) Trace un diagrama de caja de los datos.
32. La industria disquera de Estados Unidos lleva a cabo un estudio sobre el número de discos compactos de música que poseen las personas de la tercera edad y los adultos jóvenes. La información aparece en seguida.

Adultos de la tercera edad									
28	35	41	48	52	81	97	98	98	99
118	132	133	140	145	147	153	158	162	174
177	180	180	187	188					

Adultos jóvenes									
81	107	113	147	147	175	183	192	202	209
233	251	254	266	283	284	284	316	372	401
417	423	490	500	507	518	550	557	590	594

- a) Calcule la mediana y el primer y tercer cuartiles del número de compactos que poseen los ciudadanos de la tercera edad. Diseñe un diagrama de caja de la información.
b) Calcule la mediana, el primer y tercer cuartiles del número de compactos que poseen los adultos jóvenes. Diseñe un diagrama de caja de la información.
c) Compare el número de compactos que poseen ambos grupos. 
33. Las oficinas centrales de la empresa *Bank.com*, una empresa nueva de internet que realiza todas las transacciones bancarias a través de la red, se localizan en el centro de Filadelfia. El director de recursos humanos lleva a cabo un estudio relacionado con el tiempo que invierten los empleados en llegar al trabajo. La ciudad hace planes para ofrecer incentivos a las empresas que se ubiquen en el centro si estimulan a sus empleados a utilizar el transporte público. A continuación apa-

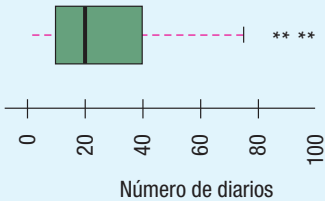
rece una lista del tiempo que se requirió esta mañana para llegar al trabajo según el empleado haya utilizado el transporte público o su automóvil.



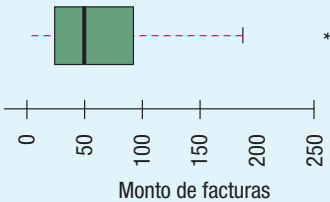
Transporte público									
23	25	25	30	31	31	32	33	35	36
37	42								

Particular									
32	32	33	34	37	37	38	38	38	39
40	44								

- a) Calcule la mediana, el primer y tercer cuartiles del tiempo de desplazamiento de los empleados utilizando el transporte público. Elabore un diagrama de caja para la información.
- b) Calcule la mediana, el primer y tercer cuartiles del tiempo de desplazamiento de los empleados en su propio vehículo. Elabore un diagrama de caja para la información.
- c) Compare los tiempos de los dos grupos.
34. El siguiente diagrama de caja muestra la cantidad de diarios que se publican en cada estado y en el Distrito de Columbia. Redacte un breve informe para resumir la cantidad que se publicó. Cerciérese de incluir información relativa a los valores del primer y tercer cuartiles, la mediana y si existe algún sesgo. Si hay datos atípicos, calcule su valor.



35. Walter Gogel Company es un proveedor industrial de cinturones de seguridad, herramientas y resortes. Las sumas de sus ingresos varían mucho, desde menos de \$20.00 hasta más de \$400.00. Durante el mes de enero enviaron 80 facturas. El siguiente es un diagrama de caja de estas facturas. Redacte un breve informe que resuma los montos de las facturas. Incluya información sobre los valores del primer y tercer cuartiles, la mediana y si existe algún sesgo. Si hay datos atípicos, aproxime el valor de estas facturas.



36. La American Society of PeriAnesthesia Nurses (ASPAN: www.aspan.org) es una organización estadounidense que agrupa a enfermeras que se desempeñan en el cuidado preanestesia y posanestesia en cirugías ambulatorias. La organización comprende 40 componentes, que se enlistan a continuación.

Estado/región	Membresía	Estado/región	Membresía
Alabama	95	Illinois	562
Arizona	399	Indiana	270
Maryland, Delaware, DC	531	Iowa	117
Connecticut	239	Kentucky	197
Florida	631	Louisiana	258
Georgia	384	Michigan	411
Hawaii	73	Massachusetts	480
(continúa)			

Estado/región	Membresía	Estado/región	Membresía
Maine	97	California	1 165
Minnesota, Dakotas	289	New Mexico	79
Missouri, Kansas	282	Pennsylvania	575
Mississippi	90	Rhode Island	53
Nebraska	115	Colorado	409
North Carolina	542	South Carolina	237
Nevada	106	Texas	1 026
New Jersey, Bermuda	517	Tennessee	167
Alaska, Idaho, Montana,		Utah	67
Oregon, Washington	708	Virginia	414
New York	891	Vermont,	
Ohio	708	New Hampshire	144
Oklahoma	171	Wisconsin	311
Arkansas	68	West Virginia	62

Utilice un software estadístico para responder las siguientes preguntas.

- Encuentre la media, la mediana y la desviación estándar del número de miembros por componente.
- Ubique el coeficiente de sesgo mediante el software. ¿Cuál es su conclusión con respecto a la forma de la distribución del tamaño del componente?
- Determine el primer y tercer cuartiles. *No* utilice el método descrito por Excel.
- Desarrolle un diagrama de caja. ¿Hay datos atípicos? ¿Cuáles componentes son atípicos? ¿Cuáles son los límites de los componentes atípicos?



37. McGivern Jewelers se ubica en Levis Square Mall, al sur de Toledo, Ohio. Recientemente publicó un anuncio en el periódico local en el que indicaba la forma, el tamaño, el precio y el grado de corte de 33 de sus diamantes en existencia. La información se muestra a continuación.

Forma	Tamaño (quilates)	Precio	Grado de corte	Forma	Tamaño (quilates)	Precio	Grado de corte
Princesa	5.03	\$44 312	Corte ideal	Redonda	0.77	\$2 828	Corte ultraideal
Redonda	2.35	20 413	Corte perfeccionado	Oval	0.76	3 808	Corte perfeccionado
Redonda	2.03	13 080	Corte ideal	Princesa	0.71	2 327	Corte perfeccionado
Redonda	1.56	13 925	Corte ideal	Talla de 58 facetas	0.71	2 732	Buen corte
Redonda	1.21	7 382	Corte ultraideal	Redonda	0.70	1 915	Corte perfeccionado
Redonda	1.21	5 154	Corte promedio	Redonda	0.66	1 885	Corte perfeccionado
Redonda	1.19	5 339	Corte perfeccionado	Redonda	0.62	1 397	Buen corte
Esmeralda	1.16	5 161	Corte ideal	Redonda	0.52	2 555	Corte perfeccionado
Redonda	1.08	8 775	Corte ultraideal	Princesa	0.51	1 337	Corte ideal
Redonda	1.02	4 282	Corte perfeccionado	Redonda	0.51	1 558	Corte perfeccionado
Redonda	1.02	6 943	Corte ideal	Redonda	0.45	1 191	Corte perfeccionado
Talla de 58 facetas	1.01	7 038	Buen corte	Princesa	0.44	1 319	Corte promedio
Princesa	1.00	4 868	Corte perfeccionado	Talla de 58 facetas	0.44	1 319	Corte perfeccionado
Redonda	0.91	5 106	Corte perfeccionado	Redonda	0.40	1 133	Corte perfeccionado
Redonda	0.90	3 921	Buen corte	Redonda	0.35	1 354	Buen corte
Redonda	0.90	3 733	Corte perfeccionado	Redonda	0.32	896	Corte perfeccionado
Redonda	0.84	2 621	Corte perfeccionado				

- Diseñe un diagrama de caja con la variable de precio y haga algún comentario sobre el resultado. ¿Hay valores atípicos? ¿Cuál es la mediana del precio? ¿Cuál es el valor del primer y tercer cuartiles?
- Diseñe un diagrama de caja de la variable de tamaño y haga comentarios sobre el resultado. ¿Hay valores atípicos? ¿Cuál es la mediana del precio? ¿Cuál es el valor del primer y tercer cuartiles?
- Diseñe un diagrama de dispersión entre las variables de precio y tamaño. Coloque el precio en el eje vertical y el tamaño en el eje horizontal. ¿Le parece que hay alguna relación entre las dos variables? ¿La relación es directa o indirecta? ¿Parece que alguno de los puntos es diferente de los demás?
- Diseñe una tabla de contingencia con las variables de forma y grado de corte. ¿Cuál es el grado de corte más común? ¿Cuál es la forma más común? ¿Cuál es la combinación más común de grado de corte y forma?



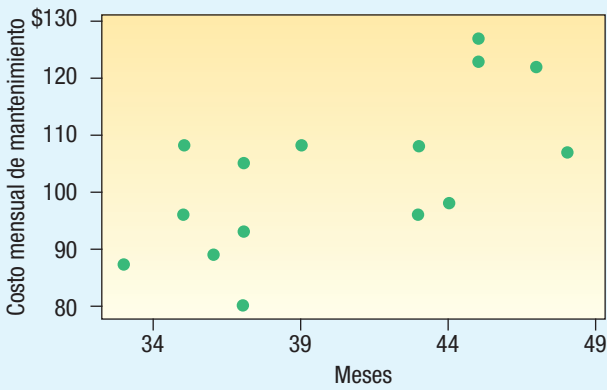
38. En la siguiente lista aparece la cantidad de comisiones que ganaron el mes pasado los ocho miembros del personal de ventas de Best Electronics. Calcule el coeficiente de sesgo utilizando ambos métodos. *Sugerencia:* El uso de una hoja de cálculo agilizará los cálculos. 

980.9	1 036.5	1 099.5	1 153.9	1 409.0	1 456.4	1 718.4	1 721.2
-------	---------	---------	---------	---------	---------	---------	---------

39. La siguiente tabla contiene la cantidad de robos de automóviles en una ciudad grande la semana pasada. Calcule el coeficiente de sesgo utilizando ambos métodos. *Sugerencia:* El uso de una hoja de cálculo agilizará las operaciones. 

3	12	13	7	8	3	8
---	----	----	---	---	---	---

40. El gerente de Servicios de Información de Wilkin Investigations, una empresa privada, estudia la relación entre el tiempo de uso (en meses) de una máquina compuesta de impresora, copiadora y fax, y el costo de mantenimiento mensual de ella. El gerente elaboró el siguiente diagrama sobre una muestra de 15 máquinas. ¿Qué puede concluir el gerente sobre la relación entre las variables?



41. Una compañía de seguros de automóvil arrojó la siguiente información relacionada con la edad de un conductor y el número de accidentes registrados el año pasado. Diseñe un diagrama de dispersión con los datos y redacte un breve resumen. 

Edad	Accidentes	Edad	Accidentes
16	4	23	0
24	2	27	1
18	5	32	1
17	4	22	3

42. Wendy’s ofrece ocho diferentes condimentos (mostaza, catsup, cebolla, mayonesa, pepinillos, lechuga, tomate y guarnición) para hamburguesas. El administrador de una de las tiendas recogió la siguiente información relativa al número de condimentos que se pidieron y el grupo de edad de los clientes. ¿Qué puede concluir respecto de la información? ¿Quién tiende a ordenar la mayor o la menor cantidad de condimentos?

Cantidad de condimentos	Edad			
	Menos de 18	De 18 a 40	De 40 a 60	60 o mayores
0	12	18	24	52
1	21	76	50	30
2	39	52	40	12
3 o más	71	87	47	28

43. La siguiente lista muestra el número de trabajadores empleados y desempleados de 20 años o mayores, de acuerdo con su género en Estados Unidos.

Género	Número de trabajadores (miles)	
	Empleados	Desempleados
Hombres	70 415	4 209
Mujeres	61 402	3 314

- ¿Cuántos trabajadores se registraron?
- ¿Qué porcentaje de trabajadores estaban desempleados?
- Compare el porcentaje de desempleados en el caso de hombres y mujeres.

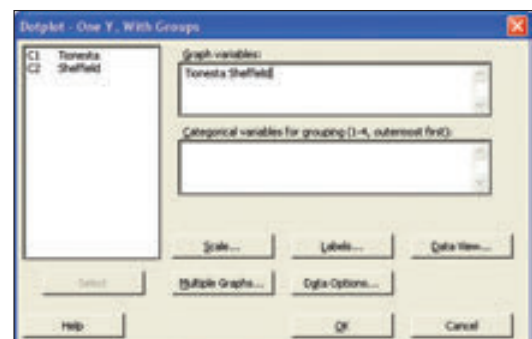
Ejercicios de la base de datos

- Consulte los datos Real Estate, que incluyen información sobre las casas vendidas en Goodyear, Arizona, el año pasado. Prepare un reporte sobre los precios de venta de las casas. Asegúrese de responder en su informe las siguientes preguntas:
 - Elabore un diagrama de caja. Estime el primer y tercer cuartiles. ¿Hay datos atípicos?
 - Desarrolle un diagrama de dispersión con el precio en el eje vertical y el tamaño de la casa en el horizontal. ¿Le parece que hay alguna relación entre las dos variables? ¿La relación es directa o inversa?
 - Elabore un diagrama de dispersión con el precio en el eje vertical y la distancia al centro de la ciudad en el horizontal. ¿Parece que hay alguna relación entre las dos variables? ¿La relación es directa o inversa?
- Busque en Baseball 2009 la información sobre los 30 mejores equipos de la Liga Mayor en la temporada 2009.
 - Seleccione la variable que se refiere al año en que el estadio fue construido. (*Sugerencia:* Reste el año en el que el estadio se construyó del año actual para determinar la edad del estadio, y trabaje con esta variable.) Diseñe un diagrama de caja ¿Hay datos atípicos?
 - Seleccione la variable relacionada con el salario del equipo y diseñe un diagrama de caja. ¿Hay datos atípicos? ¿Cuáles son los cuartiles? Redacte un breve resumen de su análisis. ¿Cómo se comparan los salarios de los Yanquis de Nueva York con los otros equipos?
 - Trace un diagrama de dispersión en cuyo eje vertical se indique el número de juegos ganados y el salario del equipo en el eje horizontal. ¿Cuáles son sus conclusiones?
 - Seleccione la variable juegos ganados. Trace un diagrama de puntos. ¿Qué conclusiones puede obtener a partir de esta gráfica?
- Consulte los datos sobre los autobuses que operan en el distrito escolar Buena.
 - Refiérase a la variable costo de mantenimiento. Desarrolle un diagrama de caja. ¿Cuáles son el primer y tercer cuartiles? ¿Hay datos atípicos?
 - Determine el costo mediano de mantenimiento. Basándose en la mediana, desarrolle una tabla de contingencias en donde el fabricante sea una variable y la otra si el costo de mantenimiento estuvo por arriba o por debajo de la mediana. ¿Cuáles son sus conclusiones?

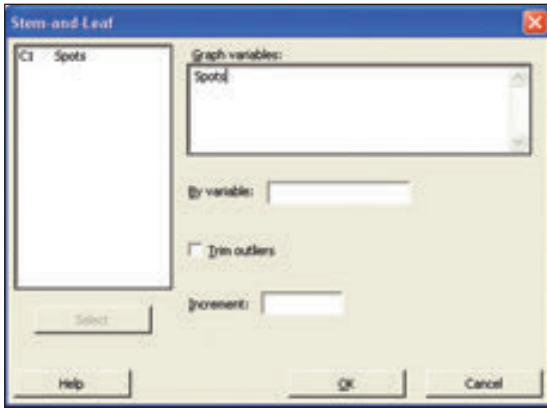
Comandos de software

- Los comandos de Minitab para elaborar el diagrama de puntos de la página 104 son los siguientes:
 - Introduzca el número de vehículos que recibieron servicio en Tionesta Ford Lincoln Mercury en la columna C1 y en Sheffield Motors en C2. Ponga el nombre adecuado a las variables.
 - Seleccione **Graph** y **Dotplot**. En el primer cuadro de diálogo, seleccione **Multiple Y's Simple** en la esquina inferior izquierda y haga clic en **OK**. En el siguiente cuadro de diálogo, seleccione **Tionesta** y **Sheffield** como variables para **Graph**, haga clic en **Labels** y escriba un título adecuado. Haga clic en **OK**.
 - Para calcular las estadísticas descriptivas que aparecen en la pantalla, seleccione **Stat**, **Basic statistics** y, en seguida, **Display Descriptive statistics**. En el cuadro de diálogo, seleccione **Tionesta** y **Sheffield** como

Variables, haga clic en **Statistics**, seleccione las estadísticas que desee obtener y, finalmente, haga doble clic en **OK**.

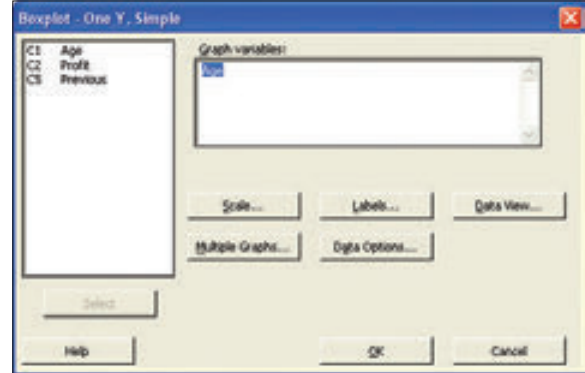


2. Los comandos de Minitab para elaborar el diagrama de tallo y hojas de la página 107 son los siguientes:
 - a) Importe los datos del sitio web del libro: www.mhhe.com/lind15e. El nombre del archivo es **Table 4-1**.
 - b) Seleccione **Graph** y haga clic en **Stem-and-Leaf**.
 - c) Seleccione la variable **Spots**, introduzca **10** como **Increment** y haga clic en seguida en **OK**.



3. Los comandos de Minitab para elaborar el resumen descriptivo de la página 113 son los siguientes:
 - a) Importe los datos de las comisiones de Smith Barney del ejemplo en la página 111.
 - b) De la barra de herramientas, seleccione **Stat, Basic Statistics** y **Display Descriptive Statistics**. En el cuadro de diálogo seleccione **Commissions** como **Variable** y en seguida haga clic en **OK**.
4. Los comandos de Excel para elaborar las estadísticas descriptivas de la página 114 son:
 - a) Ingrese los datos de las comisiones de Smith Barney del ejemplo en la página 111.
 - b) En la celda C4 escriba **Quartile 1** y en C6 escriba **Quartile 3**.
 - c) En la celda D4 escriba "**=QUARTILE(A1:A16,1)**" y presione Enter. En la celda D6 escriba "**=QUARTILE(A1:A16,3)**" y presione Enter.
5. Los comandos de Minitab para elaborar el diagrama de caja de la página 117 son los siguientes:

- a) Importe los datos de Applewood Auto Group.
- b) Seleccione **Graph** y en seguida **Boxplot**. En el cuadro de diálogo seleccione **Simple** en la esquina superior izquierda y haga clic en **OK**. Seleccione **Age** como **Graph variable**, haga clic en **Labels**, incluya un encabezado adecuado y haga clic en **OK**.



6. Los comandos de Minitab para construir el resumen descriptivo de la página 122 son los siguientes:
 - a) Ingrese los datos en la primera columna. En la celda de debajo de C1, ingrese la variable **Earnings**.
 - b) Seleccione **Stat, Basic Statistics** y haga clic en **Graphical Summary**. Seleccione **Earnings** como variable y haga clic en **OK**.
7. Los comandos de Excel para dibujar el diagrama de dispersión de la página 125 son los siguientes:
 - a) Recupere los datos de Applewood Auto Group.
 - b) Resalte con el mouse la columna de edad y la de ganancia. Incluya la primera fila.
 - c) Seleccione la pestaña **Insert**. Seleccione **Scatter** en las opciones de **Chart**. Seleccione **Chart Title** y escriba un nombre para el diagrama. Después, bajo la misma pestaña **Layout**, seleccione **AxisTitles**. En **Primary Vertical Axis Title**, escriba **Profit** como el nombre del eje. En **Primary Horizontal Axis Title**, escriba **Age** como el nombre del eje. Seleccione **Legend** y elija **None**.

Capítulo 4 Respuestas a las autoevaluaciones



- 4-1 1. a) 79, 105
 b) 15
 c) De 88 a 97; 75% de las tiendas se encuentran en este rango.

2.

7	7
8	0013488
9	1256689
10	1248
11	26

- a) 8
 b) 10.1, 10.2, 10.4, 10.8
 c) 9.5
 d) 11.6, 7.7

- 4-2 a) 7.9
 b) $Q_1 = 7.76$, $Q_3 = 8.015$

- 4-3 El valor más bajo es 10 y el más alto 85; el primer cuartil es 25 y el tercero 60. Alrededor de 50% de los valores se encuentran entre 25 y 60. El valor de la mediana es de 40. La distribución es positivamente sesgada.

4-4 a) $\bar{X} = \frac{407}{5} = 81.4$, mediana = 84

$$s = \sqrt{\frac{923.2}{5-1}} = 15.19$$

b) $sk = \frac{3(81.4 - 84.0)}{15.19} = -0.51$

c)

X	$\frac{X - \bar{X}}{s}$	$\left[\frac{X - \bar{X}}{s}\right]^3$
73	-0.5530	-0.1691
98	1.0928	1.3051
60	-1.4088	-2.7962
92	0.6978	0.3398
84	0.1712	0.0050
		-1.3154

$$sk = \frac{5}{(4)(3)} [-1.3154]$$

$$= -0.5481$$

d) La distribución es de alguna forma negativamente sesgada.

4-5 a) Diagrama de dispersión

b) 16

c) \$7 500

d) Fuerte y directa

Repaso de los capítulos 1-4

Esta sección constituye un repaso de los conceptos y términos más importantes que estructuran los capítulos 1 a 4. El capítulo 1 se inició con una descripción del significado y objetivo de la estadística. En seguida se describieron los diferentes tipos de variables y los cuatro niveles de medición. El capítulo 2 se centró en la descripción de un conjunto de observaciones y la forma en la que se organizaban en una distribución de frecuencias y en la representación de la distribución de frecuencias como un histograma o un polígono de frecuencias. El capítulo 3 comenzó con la descripción de medidas de ubicación, como la media, la media ponderada, la mediana, la media geométrica y la moda. Este capítulo también incluyó las medidas de dispersión o propagación. En esta sección se estudiaron el rango, la desviación media, la varianza y la desviación estándar. El capítulo 4 incluyó diversas técnicas de graficación, como los diagramas de puntos, los diagramas de caja y los diagramas de dispersión. También el coeficiente de sesgo, que indica la falta de simetría que puede existir en un conjunto de datos.

A lo largo de esta sección se destacó la importancia del software estadístico, como Excel y Minitab. En estos capítulos muchas capturas de pantalla demostraron la rapidez y eficacia con la que se puede organizar un conjunto de datos en una distribución de frecuencias; mostraron, asimismo, el cálculo de diversas medidas de ubicación o de variación y la información que se presenta de forma gráfica.

Glosario

Capítulo 1

Estadística Ciencia encargada de recolectar, organizar, analizar e interpretar datos numéricos con el fin de que se tomen decisiones más efectivas.

Estadística descriptiva Técnicas que se emplean para describir las características importantes de un conjunto de datos. Éstos pueden incluir la organización de los valores en una distribución de frecuencias y el cálculo debería ser de ubicación, de dispersión y sesgos.

Estadística inferencial, también denominada **inferencia estadística** Esta faceta de la estadística se relaciona con el cálculo de un parámetro basado en la estadística de una muestra. Por ejemplo, si 2 calculadoras de mano de una muestra de 10 calculadoras son defectuosas, podemos inferir que 20% de la producción es defectuosa.

Medida de intervalo Si una observación es mayor que otra por una cierta cantidad, y el punto cero es arbitrario, la medición corresponde a una escala de intervalo. Por ejemplo, la diferencia entre las temperaturas de 70 y 80 grados es de 10 grados. Asimismo, una temperatura de 90 grados es 10 grados más alta que una temperatura de 80 grados, y así sucesivamente.

Medida de razón Si las distancias entre números son de cierto tamaño constante conocido y *existe un punto cero real*, además de que la razón entre dos valores es significativa, la medida es de escala de razón. Por ejemplo, la distancia entre \$200 y \$300 es \$100, y en el caso del dinero, existe un punto cero real. Si se tienen cero dólares, no hay dinero (no se tiene nada). Asimismo, la razón entre \$200 y \$300 es significativa.

Medida nominal Nivel de medición *más bajo*. Si los datos se clasifican en categorías y el orden de dichas categorías no es

importante, se trata del nivel nominal de medición. Ejemplos de éste son el género (hombre, mujer) y la afiliación política (repblicano, demócrata, independiente, todos los demás). Si no hay diferencia entre listar primero a un hombre que a una mujer, los datos son de nivel nominal.

Medida ordinal Los datos pueden ser ordenados lógicamente refiriéndose a un orden. Por ejemplo, la respuesta del consumidor al sonido de una nueva bocina puede ser: excelente, muy buena, regular o pobre.

Muestra Porción, o subconjunto, de la población que se estudia.

Población Colección o conjunto de individuos, objetos o medidas cuyas propiedades se estudian.

Capítulo 2

Clase Intervalo en el que se recopilan los datos. Por ejemplo, \$4 a \$7 constituye una clase; \$7 a \$11 es otra.

Distribución de frecuencias Agrupación de datos en clases que muestra el número de observaciones en cada una de las clases mutuamente excluyentes. Por ejemplo, los datos se organizan en clases como las siguientes: de \$1 000 a \$2 000; de \$2 000 a \$3 000, y así sucesivamente, con el fin de resumir la información.

Distribución de frecuencias relativas Distribución de frecuencias que muestra la fracción o parte del total de observaciones de cada clase.

Exhaustivo Cada observación debe caer en alguna de las categorías.

Frecuencia de clase Número de observaciones de cada clase. Si se realizan 16 observaciones de la clase de \$4 a \$6, 16 es la frecuencia de clase.

Gráficas Formatos especiales de representación que se utilizan para mostrar una distribución de frecuencias, incluyendo histogramas, polígonos de frecuencias y polígonos de frecuencias acumulativas. Otros dispositivos gráficos que se emplean para representar datos son las gráficas de líneas, las gráficas de barras y las gráficas de pastel.

Histograma Representación gráfica de una frecuencia o una distribución de frecuencias relativas. El eje horizontal muestra las clases. La altura vertical de barras adyacentes muestra la frecuencia o frecuencia relativa de cada clase.

Mutuamente excluyente Propiedad de un conjunto de categorías que permite incluir a un individuo, objeto o medida en una sola categoría.

Punto medio Valor que divide a la clase en dos partes iguales. En las clases que van de \$10 a \$20 y de \$20 a \$30, los puntos medios son \$15 y \$25, respectivamente.

Capítulo 3

Desviación estándar Raíz cuadrada de la varianza.

Desviación media Media de las desviaciones de la media, sin tomar en cuenta los signos. Se abrevia *DM*.

Media aritmética Suma de valores dividida entre el número de valores. El símbolo de la media de una muestra es $\bar{X}$, y el símbolo de una media poblacional es μ .

Media geométrica Enésima raíz del producto de los valores. Es de particular utilidad para promediar razones de cambio y números indicadores. Minimiza la importancia de los valores extremos. Una segunda aplicación de la media geométrica se relaciona con determinar el cambio porcentual anual medio durante cierto periodo. Por ejemplo, si las ventas brutas fueron de \$245 millones en 1990 y de \$692 millones en 2010, el incremento porcentual anual promedio es 5.33.

Media ponderada Cada valor se pondera de acuerdo con su importancia relativa. Por ejemplo, si 5 camisas cuestan \$10 cada

una, y 20 cuestan \$8 cada una, el precio medio ponderado es de \$8.40: $[(5 \times \$10) + (20 \times \$8)]/25 = \$210/25 = \8.40 .

Mediana Valor de la observación media después de que todas las observaciones se ordenaron de menor a mayor. Por ejemplo, si las observaciones 6, 9 y 4 se ordenan 4, 6 y 9, la mediana es 6, el valor medio.

Medida de dispersión Valor que muestra la propagación de los datos. El rango, la varianza y la desviación estándar son medidas de dispersión.

Medida de ubicación Número que indica un solo valor típico de los datos. Señala al centro de una distribución. La media aritmética, la media ponderada, la mediana, la moda y la media geométrica son medidas de ubicación central.

Moda Valor que se presenta con mayor frecuencia en un conjunto de datos. En el caso de datos agrupados, es el *punto medio* de la clase que contiene el máximo número de valores.

Rango Medida de dispersión calculada como el valor máximo menos el valor mínimo.

Varianza. Medida de dispersión respecto de la media aritmética basada en las diferencias promedios elevadas al cuadrado.

Capítulo 4

Coefficiente de sesgo Medida de la falta de simetría de una distribución. En el caso de una distribución simétrica, no existe sesgo, así que el coeficiente de sesgo es cero. De lo contrario, puede ser positivo o negativo, con límites ± 3.0 .

Cuartiles Valores de un conjunto de datos ordenados (de mínimo a máximo) que dividen los datos en cuatro intervalos de frecuencias aproximadamente iguales.

Dato atípico Dato que suele estar muy lejos de los otros. Una regla aceptada es clasificar una observación como dato atípico si el rango intercuartil está 1.5 veces por encima del tercer cuartil o por debajo del primer cuartil.

Deciles Valores de un conjunto de datos ordenados (de mínimo a máximo), que dividen los datos en diez intervalos de frecuencias aproximadamente iguales.

Diagrama de caja Representación gráfica que muestra la forma general de la distribución de una variable. Se basa en cinco estadísticos descriptivos: los valores máximo y mínimo, el primer y tercer cuartiles y la mediana.

Diagrama de dispersión Técnica gráfica que se emplea para mostrar la relación entre dos variables medidas con escalas de intervalo o de razón.

Diagrama de puntos Herramienta de investigación que resume la distribución de una variable apilando los puntos sobre una línea de puntos que muestra los valores de la variable. Un diagrama de puntos utiliza todos los valores.

Diagrama de tallo y hojas Método para representar la distribución de una variable utilizando todos los valores. Los valores son clasificados por el dígito principal de los datos. Por ejemplo, si un conjunto de datos contiene valores entre 13 y 84, se utilizarían para los tallos ocho clases basadas en los dígitos de las decenas. Las unidades corresponderían a las hojas.

Percentiles Valores de un conjunto de datos ordenados (de mínimo a máximo) que dividen los datos en cien intervalos de frecuencias aproximadamente iguales.

Rango intercuartil Valor absoluto de la diferencia numérica entre el primer y tercer cuartiles. Cincuenta por ciento de los valores de una distribución se presentan en este rango.

Tabla de contingencia Tabla que se utiliza para clasificar observaciones de acuerdo con dos o más características nominales.

Problemas

1. Una muestra de fondos depositados en la cuenta de cheques miniatura del First Federal Savings Bank, reveló las siguientes cantidades:

\$124	\$14	\$150	\$289	\$52	\$156	\$203	\$82	\$27	\$248
39	52	103	58	136	249	110	298	251	157
186	107	142	185	75	202	119	219	156	78
116	152	206	117	52	299	58	153	219	148
145	187	165	147	158	146	185	186	149	140

Utilice un paquete de software estadístico como Excel o Minitab para ayudarse a contestar las siguientes preguntas.

- Determine la media, la mediana y la desviación estándar.
 - Determine el primer y tercer cuartiles.
 - Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.
 - Organice la distribución de fondos en una distribución de frecuencia.
 - Redacte un breve resumen de los resultados que obtuvo en los incisos a) a d).
2. A continuación se presenta una lista de los 44 presidentes de Estados Unidos y sus edades cuando comenzaron sus respectivos periodos.

Número	Nombre	Edad	Número	Nombre	Edad
1	Washington	57	23	B. Harrison	55
2	J. Adams	61	24	Cleveland	55
3	Jefferson	57	25	McKinley	54
4	Madison	57	26	T. Roosevelt	42
5	Monroe	58	27	Taft	51
6	J.Q. Adams	57	28	Wilson	56
7	Jackson	61	29	Harding	55
8	Van Buren	54	30	Coolidge	51
9	W.H. Harrison	68	31	Hoover	54
10	Tyler	51	32	F.D. Roosevelt	51
11	Polk	49	33	Truman	60
12	Taylor	64	34	Eisenhower	62
13	Fillmore	50	35	Kennedy	43
14	Pierce	48	36	L.B. Johnson	55
15	Buchanan	65	37	Nixon	56
16	Lincoln	52	38	Ford	61
17	A. Johnson	56	39	Carter	52
18	Grant	46	40	Reagan	69
19	Hayes	54	41	G.H.W. Bush	64
20	Garfield	49	42	Clinton	46
21	Arthur	50	43	G.W. Bush	54
22	Cleveland	47	44	Obama	47

Utilice un paquete de software estadístico como Excel o Minitab para ayudarse a contestar las siguientes preguntas.

- Determine la media, la mediana y la desviación estándar.
- Determine el primer y tercer cuartiles.
- Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.
- Organice la distribución de fondos en una distribución de frecuencia.
- Redacte un breve resumen de los resultados que obtuvo en los incisos a) a d).

3. Se enlista a continuación el ingreso *per cápita* de los 50 estados y el Distrito de Columbia.

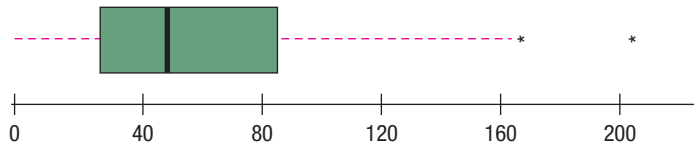
Estado	Cantidad	Estado	Cantidad
Alabama	\$30 894	Montana	\$30 790
Alaska	38 138	Nebraska	34 440
Arizona	31 936	Nevada	38 994
Arkansas	28 473	New Hampshire	39 753
California	39 626	New Jersey	46 763
Colorado	39 491	New Mexico	29 929
Connecticut	50 762	New York	44 027
Delaware	39 131	North Carolina	32 247
DC	57 746	North Dakota	32 763
Florida	36 720	Ohio	33 320
Georgia	32 095	Oklahoma	32 391
Hawaii	37 023	Oregon	33 299
Idaho	29 920	Pennsylvania	36 825
Illinois	38 409	Rhode Island	37 523
Indiana	32 288	South Carolina	29 767
Iowa	33 038	South Dakota	32 030
Kansas	34 799	Tennessee	32 172
Kentucky	29 729	Texas	35 166
Louisiana	31 821	Utah	29 406
Maine	32 095	Vermont	34 871
Maryland	43 788	Virginia	39 540
Massachusetts	46 299	Washington	38 212
Michigan	33 788	West Virginia	28 206
Minnesota	38 859	Wisconsin	34 405
Mississippi	27 028	Wyoming	40 655
Missouri	32 789		

Utilice un paquete de software estadístico como Excel o Minitab para ayudarse a contestar las siguientes preguntas.

- a) Determine la media, la mediana y la desviación estándar.
 - b) Determine el primer y tercer cuartiles.
 - c) Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.
 - d) Organice la distribución de fondos en una distribución de frecuencia.
 - e) Redacte un breve resumen de los resultados que obtuvo en los incisos a) a d).
4. Una muestra de 12 casas que se vendieron la semana pasada en St. Paul, Minnesota, reveló la siguiente información. Trace un diagrama de dispersión. ¿Es posible concluir que, conforme las dimensiones (expresadas en miles de pies cuadrados) de la casa aumentan, el precio de venta (en miles de dólares) también se incrementa?

Dimensiones de la casa (miles de pies cuadrados)	Precio de venta (miles de dólares)	Dimensiones de la casa (miles de pies cuadrados)	Precio de venta (miles de dólares)
1.4	100	1.3	110
1.3	110	0.8	85
1.2	105	1.2	105
1.1	120	0.9	75
1.4	80	1.1	70
1.0	105	1.1	95

5. Consulte el siguiente diagrama:



- ¿Cuál es el nombre de la gráfica?
- ¿Cuál es la mediana y los valores del primer y tercer cuartiles?
- ¿Es la distribución positivamente sesgada? Indique cómo lo sabe.
- ¿Hay datos atípicos? Si es el caso, estime los valores.
- ¿Puede determinar el número de observaciones en el estudio?

Casos

A. Century National Bank

El siguiente caso aparecerá en las subsecuentes secciones de repaso. Suponga que usted trabaja en el Departamento de Planeación del Century National Bank y le reporta a la señora Lamberg. Usted necesita hacer un análisis de datos y preparar un breve informe escrito. Recuerde que el señor Selig es el presidente del banco, de modo que usted querrá asegurarse de que su informe sea completo y exacto. El apéndice A.6 contiene una copia de los datos.

Century National Bank cuenta con oficinas en diversas ciudades de la región central y el sureste de Estados Unidos. Al señor Dan Selig, presidente y director ejecutivo, le gustaría conocer las características de sus clientes con cuentas de cheques. ¿Cuál es el saldo de un cliente típico?

¿Cuántos servicios bancarios más utilizan los clientes con cuentas de cheques? ¿Utilizan los clientes el servicio de cajero automático y, de ser así, cuán a menudo? ¿Qué hay de las tarjetas de débito? ¿Quién las utiliza y con cuánta frecuencia?

Para comprender mejor a los clientes, el señor Selig pidió a la señora Wendy Lamberg, directora de planeación, que seleccionara una muestra de clientes y preparara un informe. Para comenzar, ella ha nombrado un equipo de entre su personal. Usted es el jefe del equipo y el responsable de elaborar el informe. Elige una muestra aleatoria de 60 clientes. Además del saldo de cada cuenta al final del mes pasado, usted determina lo siguiente: 1) el número de transacciones en cajeros automáticos del mes pasado; 2) el número de servicios bancarios distintos (cuenta de ahorro, certificados de depósito, etc.) que utiliza el cliente; 3) si el cliente posee una tarjeta de débito (éste es un servicio bancario relativamente nuevo respecto del cual los cargos se hacen directamente a la cuenta del cliente); 4) si se paga o no interés en la cuenta de cheques. La muestra incluye clientes de las sucursales en Cincinnati, Ohio; Atlanta, Georgia; Louisville, Kentucky, y Erie, Pennsylvania.

- Diseñe una gráfica o tabla que represente los saldos de las cuentas de cheques. ¿Cuál es el saldo de un cliente típico? ¿Hay clientes con más de \$2 000 en sus cuentas? ¿Le parece que existe una diferencia en la distribución de las cuentas entre las cuatro sucursales? ¿En torno a qué valor tienden a acumularse los saldos?
- Determine la media y la mediana de los saldos de las cuentas de cheques. Compare la media y la mediana de los saldos de

las cuatro sucursales. ¿Existe alguna diferencia entre las sucursales? Explique en su informe la diferencia entre la media y la mediana.

- Determine el rango y la desviación estándar de los saldos de las cuentas de cheques. ¿Qué muestran el primer y tercer cuartiles? Determine el coeficiente de sesgo e indique lo que muestra. Como el señor Selig no maneja estadísticas diariamente, incluya una breve descripción e interpretación de la desviación estándar y de otras medidas.

B. Wildcat Plumbing Supply, Inc.: ¿hay diferencias de género?

Wildcat Plumbing Supply ha dado servicios de plomería en el sur de Arizona por más de 40 años. La compañía, que fue fundada por el señor Terrence St. Julian y hoy la dirige su hijo Cory, ha crecido de un puñado de empleados a más de 500. Cory está interesado en los diferentes puestos en la compañía en los que trabajan hombres y mujeres que llevan a cabo las mismas tareas, pero con diferente salario. Para investigar, recoge la información que sigue. Suponga que usted es un estudiante que lleva a cabo prácticas en el departamento de contabilidad y que se le ha encomendado la tarea de redactar un informe que resuma la situación.

Salario anual (miles de dólares)	Mujeres	Hombres
Menos de 30	2	0
30 a 40	3	1
40 a 50	17	4
50 a 60	17	24
60 a 70	8	21
70 a 80	3	7
80 o más	0	3

Para arrancar el proyecto, el señor Cory St. Julian organizó una junta con su personal, a la cual usted fue invitado. En esta junta se sugirió que usted calculara diversas medidas de ubicación, que trazara diagramas, como una distribución de frecuencias acumulativas y que determinara los cuartiles tanto de hom-

bres como de mujeres. Elabore los diagramas y redacte un informe que resuma los salarios anuales de los empleados de Wildcat Plumbing Supply. ¿Parece que hay diferencias de pago a partir del género?

C. Kimble Products: ¿hay alguna diferencia en el pago de comisiones?

En la junta nacional de ventas de enero, al director ejecutivo de Kimble Products se le cuestionó sobre la política de la compañía en lo que se refiere al pago de comisiones a sus representantes de ventas. La compañía vende artículos deportivos en dos mercados importantes. Tiene 40 representantes de ventas que se

comunican directamente con una gran cantidad de clientes, como los departamentos de educación física de los principales institutos, universidades y franquicias de artículos deportivos profesionales. Además, 30 agentes de ventas representan a la compañía ante tiendas de menudeo ubicadas en centros comerciales y grandes almacenes de descuento, como Kmart y Target.

Al llegar a las oficinas centrales, el director ejecutivo solicitó al gerente de ventas un informe en el que se compararan las comisiones que ganaron el año pasado las dos secciones del equipo de ventas. ¿Concluiría usted que existe alguna diferencia? En el informe incluya información sobre la tendencia central, así como sobre la dispersión en los dos grupos.

Comisiones que obtuvieron los representantes de ventas que atienden departamentos de deportes (\$)

354	87	1 676	1 187	69	3 202	680	39	1 683	1 106
883	3 140	299	2 197	175	159	1 105	434	615	149
1 168	278	579	7	357	252	1 602	2 321	4	392
416	427	1 738	526	13	1 604	249	557	635	527

Comisiones que obtuvieron los representantes de ventas que atienden grandes tiendas de menudeo (\$)

1 116	681	1 294	12	754	1 206	1 448	870	944	1 255
1 213	1 291	719	934	1 313	1 083	899	850	886	1 556
886	1 315	1 858	1 262	1 338	1 066	807	1 244	758	918

Test de práctica

Existe un cuestionario de práctica al final de cada sección de revisión, que consta de dos partes. La primera contiene diversas preguntas objetivas, por lo general con un espacio en blanco para la respuesta. La segunda consiste en problemas y ejercicios. En la mayoría de los casos, debería tomarle de 30 a 45 minutos completar el test. Los problemas requieren de una calculadora. Verifique las soluciones en la Sección de respuestas en la parte final del libro.

Parte 1: Preguntas objetivas

- La ciencia de recolectar, organizar, presentar, analizar e interpretar los datos para ayudar a tomar decisiones eficaces se llama _____. 1. _____
- Los métodos para organizar, resumir y presentar los datos de una manera informativa se llaman _____. 2. _____
- El grupo completo de individuos u objetos de interés, o las medidas que se obtienen de todos los individuos u objetos de interés se llama _____. 3. _____
- Mencione dos tipos de variables. 4. _____
- El número de habitaciones en una casa es un ejemplo de variable _____. (discreta, continua, cualitativa: elija una) 5. _____
- Los números en los jerseys de los jugadores de las Ligas Mayores de Béisbol, ¿son un ejemplo de qué nivel de medición? 6. _____
- ¿Qué ejemplo de nivel de medición sería la clasificación de estudiantes por color de ojos? 7. _____
- ¿A qué valor equivale siempre la suma de las diferencias entre cada valor y la media? 8. _____
- Un grupo de datos contiene 70 observaciones. ¿Cuántas clases sugeriría usted para construir una distribución de frecuencias? 9. _____
- ¿Qué porcentaje de los valores en un grupo de datos es siempre más grande que la mediana? 10. _____
- El cuadrado de la desviación estándar es la _____. 11. _____
- La desviación estándar asume un valor negativo cuando _____. (Todos los valores son negativos, al menos la mitad de los valores son negativos, o nunca: elija una.) 12. _____
- ¿Cuál de los siguientes es el menos afectado por un dato atípico? (media, mediana o rango: elija una) 13. _____

Parte 2: Ejercicios

- El índice de precios de valores Russell 2000 se incrementó en las siguientes cantidades los últimos tres años.

18%	4%	2%
-----	----	----

¿Cuál es la media geométrica del incremento de los tres años?

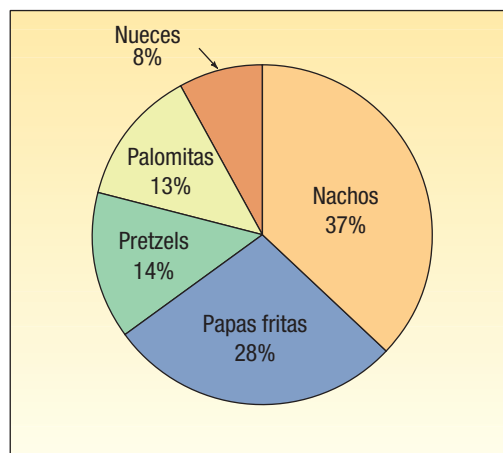
2. La siguiente información se refiere a los precios de venta, en miles de dólares, de casas que se vendieron en Warren, PA, durante 2010.

Precio de venta (miles de dólares)	Frecuencia
120.0 a 150.0	4
150.0 a 180.0	18
180.0 a 210.0	30
210.0 a 240.0	20
240.0 a 270.0	17
270.0 a 300.0	10
300.0 a 330.0	6

- a) ¿Cuál es el intervalo de clase? _____
- b) ¿Cuántas casas se vendieron en 2010? _____
- c) ¿Cuántas casas se vendieron en menos de \$210 000? _____
- d) ¿Cuál es la frecuencia relativa de la clase 210 a 240? _____
- e) ¿Cuál es el punto medio de la clase 150 a 180? _____
- f) ¿Entre cuáles dos cantidades está el rango de los precios de venta? _____
3. Una muestra de ocho estudiantes universitarios reveló que poseían el siguiente número de discos compactos.

52	76	64	79	80	74	66	69
----	----	----	----	----	----	----	----

- a) ¿Cuál es el número medio de discos compactos?
- b) ¿Cuál es el número mediano de discos compactos?
- c) ¿Cuál es el cuadragésimo percentil?
- d) ¿Cuál es el rango del número de discos compactos?
- e) ¿Cuál es la desviación estándar del número de discos compactos?
4. Un inversionista compró 200 acciones de Blair Company a 36 dólares cada una en julio de 2010, 300 acciones a 40 dólares cada una en septiembre de 2010, y 500 acciones a 50 dólares cada una en enero de 2011. ¿Cuál es la media ponderada de este inversionista del precio por acción?
5. Durante el Súper Tazón 2008 se consumieron 30 millones de libras de comida chatarra. La siguiente gráfica presenta esta información.



- a) ¿Cuál es el nombre que se le da a esta gráfica? _____
- b) Estime, en millones de libras, la cantidad de papas fritas consumidas durante el juego. _____
- c) Estime la relación entre las papas fritas y las palomitas. (El doble, la mitad, el triple, ninguna de las anteriores: elija una.) _____
- d) ¿Qué porcentaje del total comprenden las papas fritas y los nachos? _____